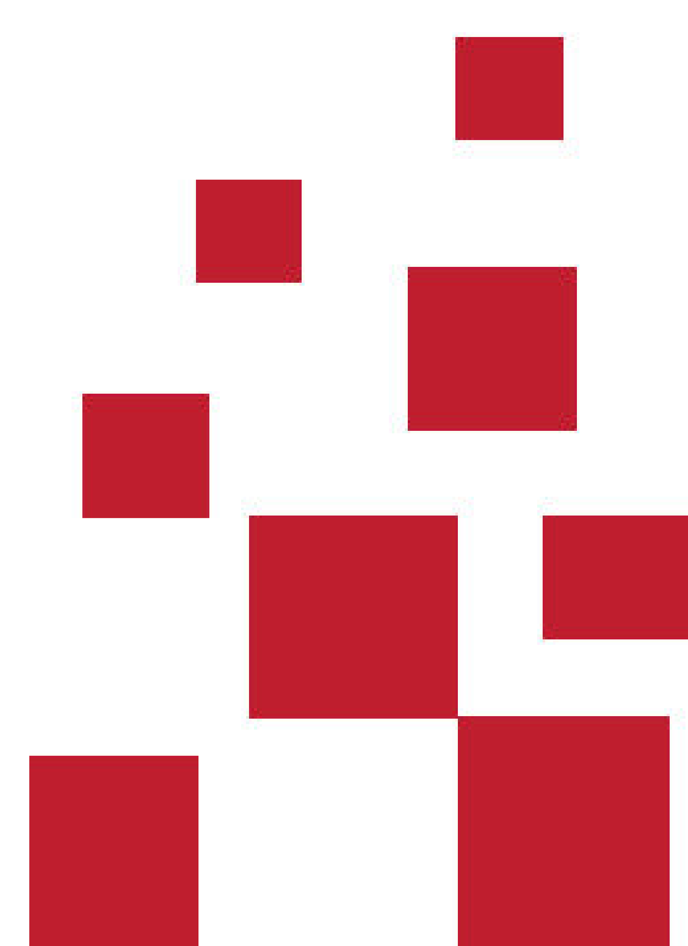


数学建模

——基于能力矩阵的自主学习新模式



王妍 斯日古冷 著



清华大学出版社

数 学 建 模

——基于能力矩阵的自主学习新模式

王　妍　斯日古冷　著

清 华 大 学 出 版 社

北　京

内 容 简 介

本书引入能力矩阵学习方法，帮助学生自主安排学习进度和控制学习难度，完成数学建模方法的学习。全书分为层次分析法、数据的处理、相关分析、回归分析、优化模型五部分，每部分包含学习情境页、能力矩阵、学习步骤计划书、学习任务页、学习信息库、教学方法和教学策略、成果展示要求、学习评价等部分。

本书适用于高职高专、贯通培养、本科及有自主学习数学建模要求的学生及教师使用。

图书在版编目(CIP)数据

数学建模：基于能力矩阵的自主学习新模式/王妍，斯日古冷著．—北京：清华大学出版社，2017
ISBN 978-7-302-44052-9

Ⅰ．①数…　Ⅱ．①王…　②斯…　Ⅲ．①数学模型　Ⅳ．①O141.4

中国版本图书馆 CIP 数据核字(2016)第 128006 号

责任编辑：刘士平
封面设计：田晓媛
责任校对：袁　芳
责任印制：宋　林

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　**邮　　编**：100084
社 总 机：010-62770175　　**邮　　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn，010-62770175-4278
印 装 者：清华大学印刷厂
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张**：15　　**字　　数**：358 千字
版　　次：2017 年 4 月第 1 版　　**印　　次**：2017 年 4 月第 1 次印刷
印　　数：1～2000
定　　价：34.00 元

产品编号：068718-01

前　言

关于数学建模，市面上有不少出版物，其中大部分是适用于本科生、研究生具有一定难度的书籍及教材，理论性较强，真正适合高职学生阅读、难易程度适宜、实用性强的书籍及教材则较少。作者总结多年数学建模教学经验、分析教学案例，以及学生的学习基础与学习习惯，结合近年来全国大学生数学建模竞赛高职赛题特点、国赛题目及所涉及的知识点，引入"能力矩阵"来帮助读者自主安排学习进度，通过小组学习、组间配合等方式，形成自主学习新模式。

本书分为复杂决策系统与层次分析法模型(12 学时)、数据的预处理(8 学时)、相关分析(4 学时)、回归分析(8 学时)、优化模型(16 学时)五部分，共 48 学时。每部分包含学习情境、能力矩阵、学习步骤计划书、学习任务页、学习信息库、教学方法和教学策略(供教师参考)、成果展示要求和学习评价几部分。

每部分以学习情境开始，由一个实际问题引入，使抽象的数学方法更加直观化，激发读者的学习兴趣。作者将每部分所需方法解构为能力矩阵。能力矩阵的每行表示该方法的学习步骤，每个学习步骤包含认知、应用及创新三个层次的能力水平，细化为 A1 至 C2 六个水平等级。若需掌握本部分的基本方法，每个学习步骤至少要达到 B1 等级。每个学习步骤配套学习任务页，顺利完成相应能力等级问题，即表示通过该等级要求。读者可根据自己(小组)的实际学习能力和进度制订学习计划(具体参见学习步骤计划书)，达到分层教学、因材施教的目的。为达到能力矩阵的各等级要求，所需部分资料可在学习信息库中查找。当然，这些资料不能涵盖全部读者所需，为达到良好的学习效果，需要在此基础上自行查阅相关信息。每部分的最后有成果展示的具体要求，以及整个学习过程(包含个人工作过程及小组工作过程)的评价与反思，以完成全部学习过程。如此，把知识点先分解，再整合，环环相扣，以学习者为中心，最大限度地降低读者的学习难度，提高其学习兴趣，也不失知识的系统性学习。

本书有助于通过学生自主设定学习计划及成果汇报任务，提高学生自主学习、创新实践的能力；有助于通过对具体情况的训练，提高学生解决实际问题的能力，提高计算机应用水平，进而提高数学实践能力；有助于通过完成团队任务，提高学生团队协作与沟通表达能力。

本书依托德国奥斯卡-冯-米勒学校最先提出的"能力矩阵教学设计方法"，感谢 Howe 女士和 Schmolders 博士向作者介绍此方法，感谢黄海洋教授、田志英教授对书稿的意见与建议，感谢北京市教委及北京财贸职业学院对作者的大力支持。本书写作历经四年时间，参阅、借鉴了大量资料，绝大部分案例经过反复推敲、改良、修订，以适应高职学生需求。应该说，没有各位亲朋、老师的帮助，本书不可能顺利出版，现一并致谢。由于作者水平有限，书中难免有错漏之处，敬请读者批评指正。

作　者

2016 年 8 月

目　　录

第一部分

复杂决策系统与层次分析法模型

学习情境——交通设计方案决策

北京市有一家商场，由于所处街道狭窄，经常造成交通堵塞，如图 1-1 所示。市政府决定解决这个问题，以改善交通环境。

经过专家组调查研究，制定出三个可行方案。

P1：在商场附近修建一座环形天桥，如图 1-2 所示。



图 1-1　商场及街道



图 1-2　环形天桥

P2：在商场附近修建地下人行通道，如图 1-3 所示。

P3：搬迁商场。

根据当地实际情况，专家组拟定五个评价准则。

C1：通车能力。

C2：方便群众。

C3：基建费用不要太高。

C4：交通安全。

C5：市容美观。

请你对改善交通环境问题提出决策建议。



图 1-3　地下人行通道

能力矩阵——层次分析法模型

学生活动	教师控制/教师组织			自我导向/自我组织		
教师角色	讲授、辅助咨询			教练		
学习步骤	能力水平					
	认知		应用		创新	
	A1	A2	B1	B2	C1	C2
	识记	理解	应用	分析	创造	评价
1. 层次分析法的思想、步骤	能记住层次分析法适用于解决问题的类型及解决流程	能理解某实际案例是否可以利用层次分析法解决	能应用层次分析法将人的思维过程层次化、数量化，为复杂决策问题提供解决途径	能分析层次分析法是定性与定量相结合的多准则决策	能体会层次分析法步骤中分解、判断、综合的系统思维方式的创新性	能评价出层次分析法的适用范围和不足之处
2. 构造层次分析结构	能记住目标层、准则层、方案层的含义	能理解将复杂决策问题进行条理化、层次化的逻辑关系	能建立并画出层次分析结构，明确问题	能在对问题的全面、深入认识的基础上，分析问题并建立层次结构	能结合指标体系建立标准，创建层次结构	能对所创建层次结构给出合理的说明和评价
3. 构造因素判断矩阵	能记住判断矩阵的构造形式及其元素性质	能理解九级标度法	能应用九级标度法给出因素判断矩阵	能根据九级标度法给出判断矩阵的合理解释	能针对实际问题创造出合理的调查问卷来调查判断矩阵	能根据专家调查问卷给出因素判别矩阵
4. 检验判断矩阵的一致性	能记住 CI、RI、CR 的含义	能理解判断思维一致性的目的和意义	能应用现有公式进行一致性检验			
5. 层次单排序	能记住何为层次单排序	能理解层次单排序与重要性顺序权重的联系，理解如何求层次单排序	能应用公式，明白单排序的计算方法	能针对排序结果进行合理分析		
6. 层次总排序	能记住何为层次总排序	能理解总排序一致性检验方法	能应用现有公式计算层次总排序	能针对总排序给出最终决策评价方案		
7. 层次分析法模型的MATLAB软件实现	能记住矩阵输入、矩阵计算等初级语法	能理解 eig 命令	能应用 MATLAB 软件求解层次分析法问题	能对软件运行结果进行分析处理，得出最终答案	能创造出层次分析法的 M 文件	能利用软件写出层次分析法的 GUI 用户图形界面

学习步骤计划书

姓　　名		班　　级	
学习日期		汇报日期	
学习任务			

学习目标	
(1) 我的目标是什么? **注意**:每一个目标都应该以"我能……"开始,并将每一个目标编号;请注明通过哪一个学习任务达到该学习目标。 (2) 我想要学会什么? **注意**:要给出在哪个时间点达到这个目标。	

学习途径	
(1) 我通过哪些途径获得该目标相关信息? (2) 我需要什么材料? (3) 我可以通过哪些方式达到该目标? (4) 达到此目标的过程中会出现什么障碍? (5) 在完成此目标的每一天中,我要完成什么?	

学习证明	
(1) 通过何种展示方式证明我达到这个目标? 如论文写作、PPT 展示、流程图绘制等。 (2) 怎样检测我是否达到这个目标? (3) 请给出预计展示成果的时间。	

学习评价	
(1) 对于学习结果,我是否满意? (2) 我通过什么方法评价自己的学习质量? (3) 我的学习信心是否充足? 我如何克服上面提到的障碍? (4) 我还能做些什么,来完善学习成果?	

学习任务页

学习任务页-1　层次分析法的思想、步骤

学生信息__________　　任务开始日期__________　　任务分值___10 分___

1-A1-1　层次分析法可以解决什么类型的问题？

1-A1-2　层次分析法的步骤、流程是什么？

1-A2　对于下列两个问题，可否应用层次分析法解决？请分别说明。

(1) 大学毕业生在就业时“双向选择”，用人单位与毕业生有各自的选择标准和要求。就毕业生来说，选择单位的标准和要求是多方面的，例如：

① 能发挥自己的才干，为国家做出较好贡献(即工作岗位适合发挥专长)。

② 工作收入较好(待遇好)。

③ 生活环境好(大城市、气候等工作条件)。

④ 单位名声好(声誉)。

⑤ 工作环境好(人际关系和谐等)。

⑥ 发展、晋升机会多(如新单位发展有后劲)等。

当你毕业时，如有多个用人单位供你选择，你将如何做出合理的决策和选择？

(2) 2010 年全国大学生数学建模竞赛 D 题。

学生宿舍事关学生在校期间的生活品质，直接或间接地影响到学生的生活、学习和健康成长。学生宿舍的使用面积、布局和设施配置等设计既要让学生生活舒适，也要方便管理，同时要考虑成本和收费的平衡，还与所在城市的地域、区位、文化习俗和经济发展水平有关。因此，学生宿舍的设计必须考虑经济性、舒适性和安全性等问题。

① 经济性：建设成本、运行成本和收费标准等。

② 舒适性：人均面积、使用方便、互不干扰、采光和通风等。

③ 安全性：人员疏散和防盗等。

请针对宿舍建设图纸，根据其经济性、舒适性和安全性，做出综合量化评价和比较。

1-B1　层次分析法的思想原理是什么?

1-B2　在层次分析法中，如何将人的判断、比较进行定量描述?

1-C1　请讨论下述建筑工程评标问题，给出解决问题的大体思路。

工程建设招投标是国际上通用的工程建设承包方式，是国际公认的选择施工企业的有效方法。

对于某工程项目，按工程基本要求和评价体系特点选择工程报价(万元)、施工方案合理性、施工工期(天)、施工单位质量实绩、企业信誉度、项目经理等级等6项指标作为评价参数。其中，定性指标评定等级由招标评委综合评定。工期预计268天。共有4家单位参加该项目的施工投标。4家投标单位的基本情况如表1-1所示。

表1-1　投标方案评价指标汇总表

单位	报价	工期	质量实绩	施工方案	企业信誉	项目经理
甲	920	230	优	优	好	一级
乙	916	246	良	优	很好	二级
丙	911	268	优	良	好	二级
丁	903	255	中	良	一般	一级

1-C2　请写出层次分析法的优缺点，并进行分析。

学习任务页-2　构造层次分析结构

学生信息__________　　任务开始日期__________　　任务分值___15 分___

2-A1　层次分析模型主要分几层？分别是什么？请用你自己的话解释，并举实际例子说明。

2-A2　每层元素一般不超过多少个？为什么？

2-B1-1　请建立下述问题的层次结构。

若你想在“黄金周”去旅游，目前需在泰山、杭州和承德三处选择一个旅游点。要考虑景点的景色、居住环境、饮食特色、交通状况和旅游费用等。请利用层次分析法进行分析、决策。

2-B1-2　请建立下述问题的层次结构。

一年一度的全国大学生数学建模竞赛是全国大学生四大竞赛之一，在全国各高校中都受到高度重视。由于竞赛场地、经费等原因，并不是所有希望参加竞赛的学生都能被录取。为了选拔出优秀的学生组队代表学校参加全国竞赛，数学建模教练组需要投入大量的精力。现在请你们用数学建模的方法来帮帮忙。

如果想在全国数学建模竞赛中取得优异的成绩，需要建模团队各成员具有一定的数学基础和必要的数学建模知识，具备熟练使用数学软件的能力，较强的语言表达和写作能力，良好的团队合作精神，还要求思维敏捷，对建立数学模型有较好的悟性等方面的素质。当然，这些能力要求中重要性有所不同。

目前选拔团队成员主要考虑以下几个因素：数学建模培训课程的出勤记录；数学建模的课堂练习成绩；上机操作成绩；学生个人简介；面试成绩；老师和学生的推荐等。

2-B1-3　请建立下述问题的层次结构。

随着我国企业兼并增多，产权关系逐步理顺，兼并市场的发育不断完善，企业兼并的竞争度和选择性大大增强。此时，生产经营好的企业如何找到最合适的扩张对象，将是一个理智的、科学的、严密的分析过程。企业兼并目标选择决策是企业管理中具有决定性意义的工作。

根据我国企业兼并现状和发展趋势的分析，企业兼并备选目标的评价主要包括以下6个方面：财务经济状况(F_1)、产品市场需求状况(F_2)、发展环境(F_3)、技术进步潜力(F_4)、组织管理情况(F_5)、工艺—技术相关性(F_6)。这6个因素直接关系到企业兼并类型的确定和方案实施。

下面请以存在3个兼并备选企业A、B、C的情形进行讨论。

2-B1-4　请建立下述问题的层次结构。

风险投资是指通过一定的机构和方式向各类机构或个人筹集风险资本，然后将其投入具有高度不确定性的中小型高新技术企业或项目，并以一定的方式参与所投资风险企业或项目的管理，期望通过实现的高成长率并最终通过出售股权等方式获得高额中长期收益的一种投资体系。

风险投资多以投资基金方式运作。运作程序包括项目评估与决策、谈判和签订协议、辅导管理及退出四个阶段，其中以项目评估与决策最关键，因为它是整个风险投资程序中至关重要的第一阶段，也是提高风险资金利用效率和决定风险投资项目成功与否的关键。

风险投资项目的不确定性主要表现在项目的技术是否具有超前意识，是否可以实现，投资产品是否具有广阔的市场前景，市场占有率会有多大，产品的市场竞争能力如何等方面。

根据对风险投资因素的分析，各个备选项目的评价主要包括以下几个方面。

(1) 技术(F_1)：指技术开发方面的各种不确定因素，如技术难度、技术适用性、技术成熟性、技术配套性、技术生命周期等。

(2) 市场潜力(F_2)：指难以确定的市场需求、产品竞争力、上市时机、市场扩展速度、潜在竞争者影响、产品替代性等。

(3) 管理(F_3)：即人员素质与经验、领导判断与决策的科学化、企业组织合理性、项目管理机制等。

(4) 领导者素质(F_4)：指创业者的专业知识水平、领导水平、能力、性格等。

请根据以上描述，确定风险投资项目选择的层次结构模型。

2-B2-1　请利用 Office 软件的编辑器画出“任务 2-B1-1”的层次结构模型。

2-B2-2　请利用 Office 软件的编辑器画出“任务 2-B1-2”的层次结构模型。

2-B2-3　请利用 Office 软件的编辑器画出“任务 2-B1-3”的层次结构模型。

2-B2-4　请利用 Office 软件的编辑器画出“任务 2-B1-4”的层次结构模型。

2-C1　请自行查阅背景资料，找出下述问题的影响因素指标。

2008 年全国大学生数学建模竞赛 D 题(部分)：NBA 赛程的分析与评价。

NBA 是全世界篮球迷们最钟爱的赛事之一，姚、易加盟以后，更是让中国球迷宠爱有加。对于 NBA 这样庞大的赛事，编制一个完整的、对各球队尽可能公平的赛程是一件非常复杂的事情，赛程的安排对球队实力的发挥和战绩有一定的影响。从报刊上经常看到球员、教练和媒体对赛程的抱怨或评论。

为了分析赛程对某一支球队的利弊，你认为有哪些要考虑的因素。

2-C2　请自行查阅综合评价模型影响因素指标的设定标准，并对“任务 2-C1”的影响指标进行说明与评价。



学习任务页-3　构造因素判断矩阵

学生信息__________　　任务开始日期__________　　任务分值____20 分____

3-A1-1　因素判断矩阵的构造原理是什么？

3-A1-2　对 n 个影响元素来说，若设两两比较判断矩阵为 $\boldsymbol{C}=(c_{ij})_{n\times n}$，则其中的 c_{ij} 表示什么？

3-A1-3　判断矩阵 $\boldsymbol{C}$ 的元素 c_{ij} 有何性质特征？

3-A2　基于九级标度法的矩阵标度及其含义分别是什么？

3-B1-1　对于“任务 2-B1-1”，请根据你的实际情况，给出判断矩阵。

3-B1-2　对于“任务 3-B1-1”，请利用 Mathtype(公式编辑器)进行排版。

3-B2　对于“任务 3-B1-1”，请给出文字解释和描述。

3-C1　对于较复杂的问题，如“任务 2-B1-2”“任务 2-B1-3”“任务 2-B1-4”等，判断矩阵是经过多位专家填写咨询表或者调查问卷之后形成的。请针对“任务 2-B1-2”设计调查问卷，并对相关专家(如数学建模指导教师等)进行调查。

3-C2　根据调查问卷(参见“任务 3-C1”)，给出“任务 2-B1-2”的因素判断矩阵。



学习任务页-4　检验判断矩阵的一致性

学生信息__________　　任务开始日期__________　　任务分值　15 分

4-A1-1　指标 CI 是什么？如何计算？

4-A1-2　指标 RI 是什么？又该如何计算？

4-A1-3　如何判定因素判断矩阵的一致性？指标 CR 如何计算？

4-A2　一致性检验的目的和意义是什么？

4-B1　对"任务 3-B1-1"所给出的因素判断矩阵进行一致性判定。

学习任务页-5　层次单排序

学生信息＿＿＿＿＿　任务开始日期＿＿＿＿＿　任务分值＿20 分＿

5-A1　何谓层次单排序？

5-A2　如何求层次单排序？

5-B1　针对“任务 2-B1-1”，求此任务的层次单排序。

5-B2　根据“任务 5-B1”的结论，给出层次单排序的结果分析。



学习任务页-6　层次总排序

学生信息__________　　任务开始日期__________　　任务分值___10 分___

6-A1　何谓层次总排序？

6-A2　总的一致性检验如何操作？

6-B1-1　计算"任务 2-B1-1"的总排序。

6-B1-2　计算“任务 2-B1-2”的总排序。

6-B2-1　根据“任务 6-B1-1”的总排序，对最终评价结果给出文字解释。

6-B2-2　根据“任务 6-B1-2”的总排序，对最终评价结果给出文字解释。



学习任务页-7　层次分析法模型的 MATLAB 软件实现

学生信息＿＿＿＿＿　　任务开始日期＿＿＿＿＿　　任务分值＿10 分＿

7-A1　针对“任务 2-B1-1”，在 MATLAB 软件中输入因素判断矩阵(参见“任务 3-B1-1”)。

7-A2　如何利用 MATLAB 软件计算矩阵的特征值及特征向量？命令的语法如何？

7-B1　针对“任务 2-B1-1”，利用 MATLAB 软件求解。

7-B2　针对“任务 7-B1”中计算出的结果进行分析处理，得出最终结果。

7-C1　利用 MATLAB 软件，创建 M 文件，以便利用软件求解。

7-C2　利用 MATLAB 软件，编写层次分析法的 GUI 用户图形界面，以便利用软件求解。



学习任务页-zh1　物流中心选址决策评价

学生信息＿＿＿＿＿　　任务开始日期＿＿＿＿＿　　任务分值＿25 分＿

物流中心是处于枢纽或重要位置，具有完整的物流环节，并能将物流集散、信息处理和控制等功能实现一体化运作的物流节点。

现归纳影响物流中心的各个因素，其层次结构如图 1-4 所示。

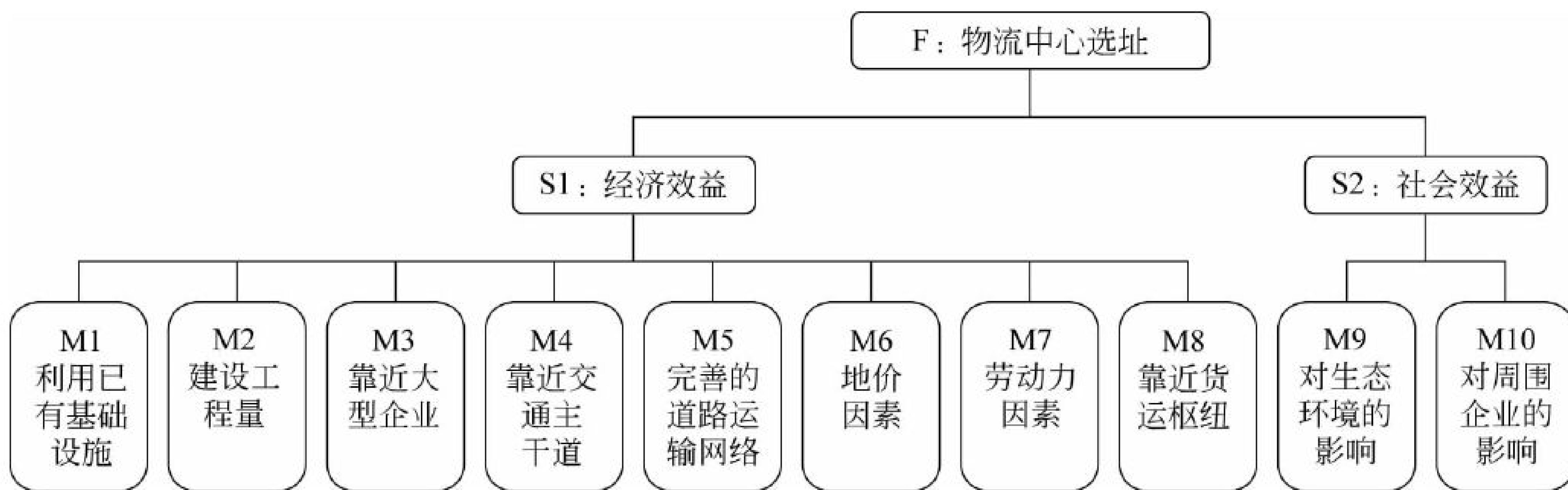


图 1-4　物流中心选址层次结构图

请采用层次分析法进行物流中心选址，即结合物流中心的职能及选址原则，考虑经济效益和社会效益，寻求最佳的选址评价方案。

学习任务页-zh2　对学生宿舍设计方案的评价

学生信息__________　　任务开始日期__________　　任务分值　25 分

2010 年全国大学生数学建模竞赛 D 题：对学生宿舍设计方案的评价。

学生宿舍事关学生在校期间的生活品质，直接或间接地影响到学生的生活、学习和健康成长。学生宿舍的使用面积、布局和设施配置等的设计，既要让学生生活舒适，也要方便管理，同时要考虑成本和收费的平衡，还与所在城市的地域、区位、文化习俗和经济发展水平有关。因此，学生宿舍的设计必须考虑经济性、舒适性和安全性等问题。

经济性：建设成本、运行成本和收费标准等。

舒适性：人均面积、使用方便、互不干扰、采光和通风等。

安全性：人员疏散和防盗等。

图 1-5～图 1-8 是四种比较典型的学生宿舍设计方案。请用数学建模的方法，就它们的经济性、舒适性和安全性做出综合量化评价和比较。



图 1-5　学生宿舍设计方案 1

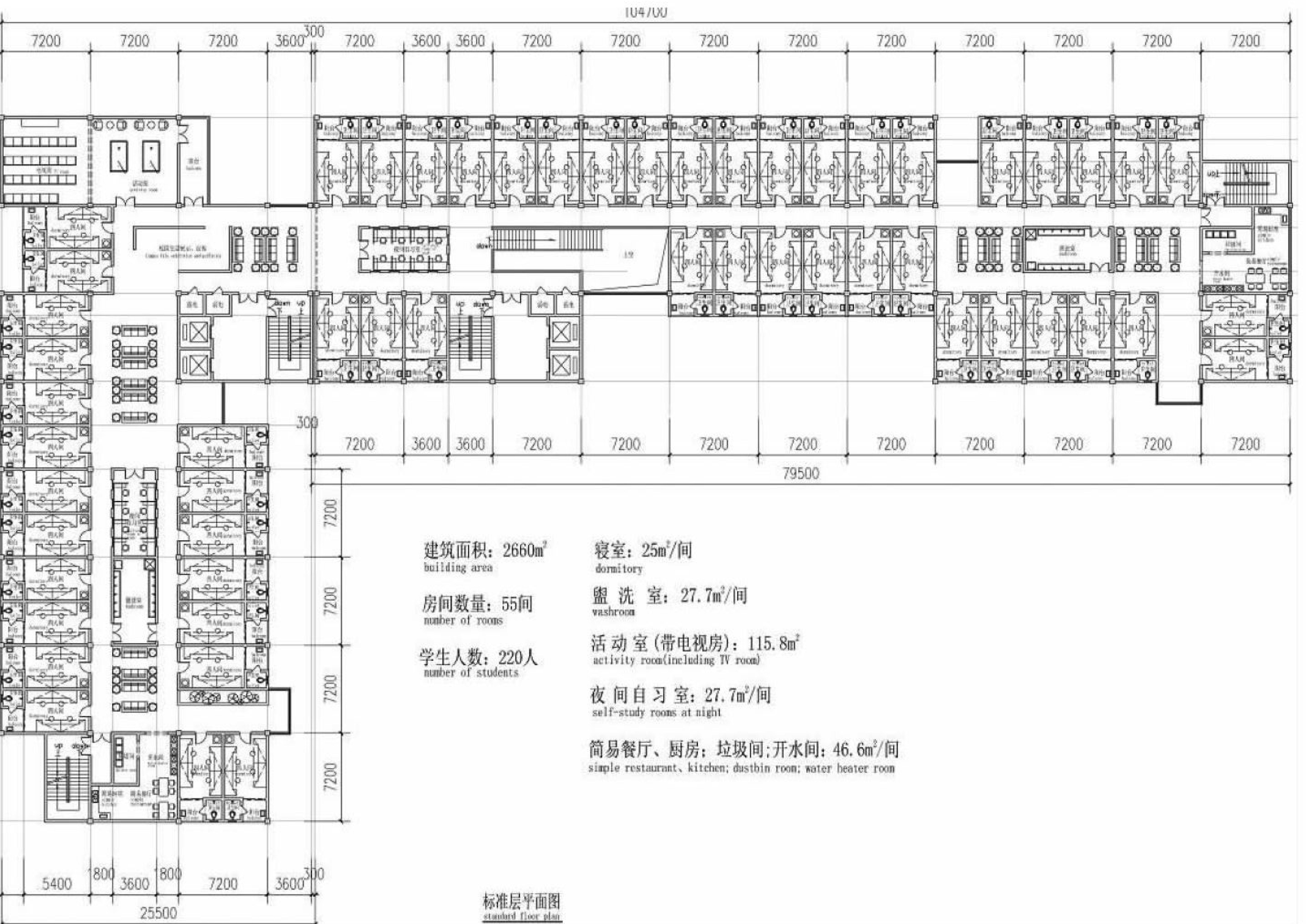


图 1-6　学生宿舍设计方案 2



图 1-7　学生宿舍设计方案 3

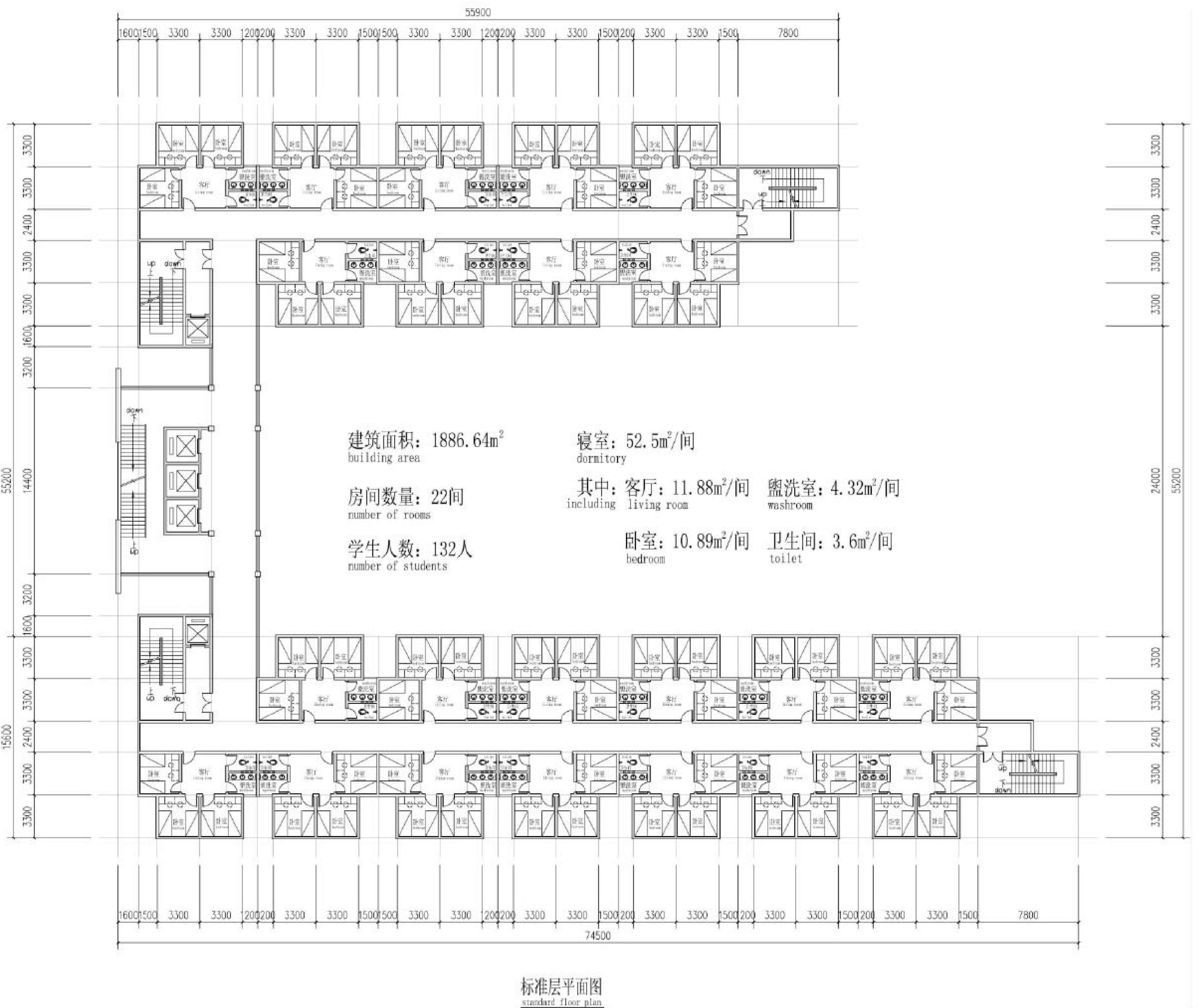


图 1-8 学生宿舍设计方案 4

学习信息库 1——层次分析法概述

综合评价方法是指利用多项指标对某个评价对象的某种属性进行定性、定量评估，或者对多个评价对象的属性进行定性、定量评估，对优劣程度排序。层次分析法（Analytic Hierarchy Process，AHP）是综合评价方法中常用的一种，是对较为复杂、较为模糊的问题做出决策的简易方法，特别适用于难以完全定量分析的复杂系统问题。它是美国运筹学家 T. L. Satty 教授于 20 世纪 70 年代初期提出的一种简便、灵活而又实用的多准则决策方法。

1. 层次分析法的基本原理与操作步骤

人们在进行社会问题、经济问题以及科学管理领域问题的系统分析中，面临的常常是一个由相互关联、相互制约的众多因素构成的复杂，但缺少定量数据的系统。层次分析法为这类问题的决策提供一种新的、简洁而实用的建模方法。

运用层次分析法建模，大体上分为以下四个步骤。

（1）建立层次结构模型

应用层次分析法分析、决策问题时，首先要把问题条理化、层次化，构造出一个有层次的结构模型。在这个模型下，复杂问题被分解为元素等组成部分。这些元素又按其属性及关系形成若干层次。上一层次的元素作为准则，对下一层次有关元素起支配作用。这些层次分为以下三类。

① 最高层：这一层次中只有一个元素，一般它是分析问题的预定目标或理想结果，因此也称为目标层。

② 中间层：这一层次中包含为实现目标所涉及的中间环节。它可以由若干层次组成，包括需要考虑的准则、子准则，因此也称为准则层。

③ 最低层：这一层次包括为实现目标可供选择的各种措施、决策方案等，因此也称为措施层或方案层。

层次结构中的层次数与问题的复杂程度及需要分析的详尽程度有关。一般情况下，层次数不受限制。每一层次中各元素所支配的元素一般不超过 9 个。这是因为支配的元素过多，会给两两比较、判断带来困难。

下面结合一个实例说明层次结构的建立。

例 1-1 若你想在“黄金周”假期去旅游，目前有杭州、泰山、承德 3 个旅游胜地供选择。试根据实际情况确定一个最佳地点。

此时，你会根据费用、景色、住宿、饮食和交通等一些准则去比较 3 个候选地点，建立如图 1-9 所示的层次结构模型。

（2）构造因素判断矩阵

层次结构反映因素之间的关系，但准则层中的各准则在目标衡量中所占的比重并不一定相同，在决策者的心中，它们各占有一定的比例。

在确定影响某因素的诸因子所占的比重时，遇到的主要困难是这些比重通常不易量化。此外，当影响某因素的因子较多时，直接考虑各因子对该因素有多大程度的影响时，常会因考虑不周全、顾此失彼而使决策者提出与他实际认为的重要程度不相一致的数据，甚至有可

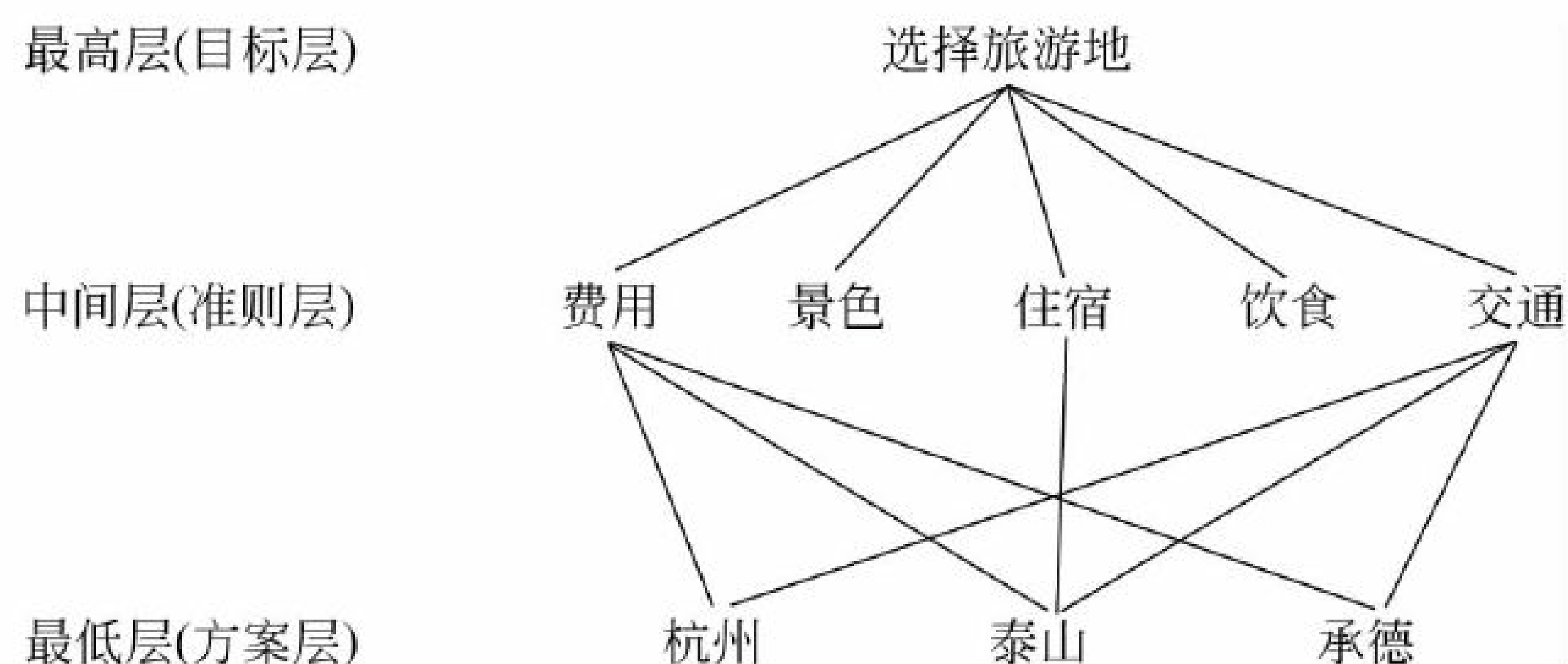


图 1-9　层次结构模型

能提出一组隐含矛盾的数据。

现在要比较 n 个因子 $X=\{x_1,x_2,\cdots,x_n\}$ 对某因素 Z 的影响大小，怎样比较才能提供可信的数据呢？Satty 等人建议，采取对因子两两比较后建立矩阵的办法。即每次取两个因子 x_i 和 x_j，以 a_{ij} 表示 x_i 和 x_j 对 Z 的影响大小之比，全部比较结果用矩阵 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 表示，称 $\boldsymbol{A}$ 为 Z 与 X 之间的因素判断矩阵(简称判断矩阵)。容易看出，若 x_i 与 x_j 对 Z 的影响之比为 a_{ij}，则 x_j 与 x_i 对 Z 的影响之比应为 $a_{ji}=\dfrac{1}{a_{ij}}$。

关于如何确定 a_{ij} 的值，Satty 等人建议，引用数字 1～9 及其倒数作为标度，此法称为九级标度法。从心理学观点来看，分级太多，会超越人们的判断能力，既增加作判断的难度，又容易因此而产生互相矛盾的数据。Satty 等人还用实验方法比较了在各种不同标度下，人们判断结果的正确性。实验结果表明，采用九级标度最合适。表 1-2 中列出了九级标度的含义。

表 1-2　九级标度的含义

标度	含　　义
1	表示两个因素相比，具有相同的重要性
3	表示两个因素相比，前者比后者稍重要
5	表示两个因素相比，前者比后者明显重要
7	表示两个因素相比，前者比后者强烈重要
9	表示两个因素相比，前者比后者极端重要
2,4,6,8	表示上述相邻判断的中间值
倒数	若因素 i 与因素 j 的重要性之比为 a_{ij}，那么因素 j 与因素 i 重要性之比为 $a_{ji}=\dfrac{1}{a_{ij}}$

(3) 层次单排序及一致性检验

判断矩阵 $\boldsymbol{A}$ 对应于最大特征值 $\lambda_{\max}$ 的特征向量 $\boldsymbol{W}$，经归一化后，即为同一层次相应因素对于上一层次某因素相对重要性的排序权值，这一过程称为层次单排序，即单排序权重为因素判断矩阵最大特征值所对应的归一化的特征向量。

上述构造成对比较判断矩阵的办法虽能减少其他因素的干扰，较客观地反映出一对因子影响力的差别，但综合全部比较结果时，其中难免包含一定程度的非一致性。如果比较结果是前后完全一致的，则矩阵 $\boldsymbol{A}$ 的元素还应当满足：

$$a_{ij}a_{jk}=a_{ik}\quad(\forall\, i,j,k=1,2,\cdots,n)\tag{1-1}$$

可以证明，矩阵 $\boldsymbol{A}$ 的最大特征值 $\lambda_{\max}$ 是否等于矩阵的阶数 n，可以用来检验是否具有一

致性。由于特征值连续地依赖于 a_{ij}，故 $\lambda_{\max}$ 比 n 大得越多，$\boldsymbol{A}$ 的非一致程度越严重，$\lambda_{\max}$ 对应的标准化特征向量越不能真实地反映出 $X=\{x_1,x_2,\cdots,x_n\}$ 在对因素 Z 的影响中所占的比重。因此，对决策者提供的判断矩阵有必要做一次一致性检验，以决定是否能接受它。

判断矩阵的一致性检验步骤如下所述。

① 计算一致性指标 CI：

$$CI=\frac{\lambda_{\max}-n}{n-1} \tag{1-2}$$

② 查找相应的平均随机一致性指标 RI。

Satty 用随机方法构造 500 个样本矩阵：随机地从 1～9 及其倒数中抽取数字构造正互反矩阵，求得最大特征根的平均值 $\bar{\lambda}_{\max}$，并定义 $RI=\dfrac{\bar{\lambda}_{\max}-n}{n-1}$。对于 $n=1,2,\cdots,9$，RI 如表 1-3 所示。

表 1-3　平均随机一致性指标 *RI* 的值

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

③ 计算一致性比例 CR：

$$CR=\frac{CI}{RI} \tag{1-3}$$

当 $CR<0.1$ 时，认为判断矩阵的一致性是可以接受的，否则应对因素判断矩阵做适当修正。

(4) 层次总排序及一致性检验

上面得到的是一组元素对其上一层中某元素的权重向量。最终要得到各元素，特别是最低层中各方案对于目标的排序权重，从而进行方案选择。总排序权重需要自上而下地将单准则下的权重进行合成。

设上一层次(A 层)包含 $A_1,A_2,\cdots,A_m$ 共 m 个因素，它们的层次总排序权重分别为 $a_1,a_2,\cdots,a_m$。又设其后的下一层次(B 层)包含 n 个因素 $B_1,B_2,\cdots,B_n$，它们关于 A_j 的层次单排序权重分别为 $b_{1j},b_{2j},\cdots,b_{nj}$(当 B_i 与 A_j 无关联时，$b_{ij}=0$)。现求 B 层中各因素关于总目标的权重，即求 B 层各因素的层次总排序权重 $b_1,b_2,\cdots,b_n$。按表 1-4 所示方式计算，即 $b_i=\sum_{j=1}^{m}b_{ij}a_j\,(i=1,2,\cdots,n)$。

表 1-4　*B* 层各因素的层次总排序权重

层 A / 层 B	A_1 a_1	A_2 a_2	… …	A_m a_m	B 层总排序权值
B_1	b_{11}	b_{12}	…	b_{1m}	$\sum_{j=1}^{m}b_{1j}a_j$
B_2	b_{21}	b_{22}	…	b_{2m}	$\sum_{j=1}^{m}b_{2j}a_j$
⋮	⋮	⋮	⋮	⋮	⋮
B_n	b_{n1}	b_{n2}	…	b_{nm}	$\sum_{j=1}^{m}b_{nj}a_j$

对层次总排序也需做一致性检验，检验仍像层次总排序那样，由高层到低层逐层进行。这是因为虽然各层次均已经过层次单排序的一致性检验，各层对比较判断矩阵都已具有较满意的一致性，但当综合考察时，各层次的非一致性仍有可能积累起来，引起最终分析结果较严重的非一致性。

设 B 层中与 A_j 相关的因素的成对比较判断矩阵在单排序中经一致性检验，求得单排序一致性指标为 $CI_j(j=1,2,\cdots,m)$，相应的平均随机一致性指标为 RI_j（CI_j、RI_j 已在层次单排序时求得），则 B 层总排序随机一致性比例为

$$CR=\frac{\sum_{j=1}^{m} CI_j a_j}{\sum_{j=1}^{m} RI_j a_j} \tag{1-4}$$

当 $CR<0.1$ 时，认为层次总排序结果具有较满意的一致性，接受该分析结果。

2. 层次分析法的应用

经过几十年的发展，许多学者改进和完善了 AHP 的缺点，形成了一些新理论和新方法，如群组决策、模糊决策和反馈系统理论近几年成为该领域的新热点。学有余力的读者可以查阅相关资料。

下面将完整分析例 1-1，完成整个层次分析法建模过程。

例 1-1（续） 为了描述方便，将各元素命名如图 1-10 所示。

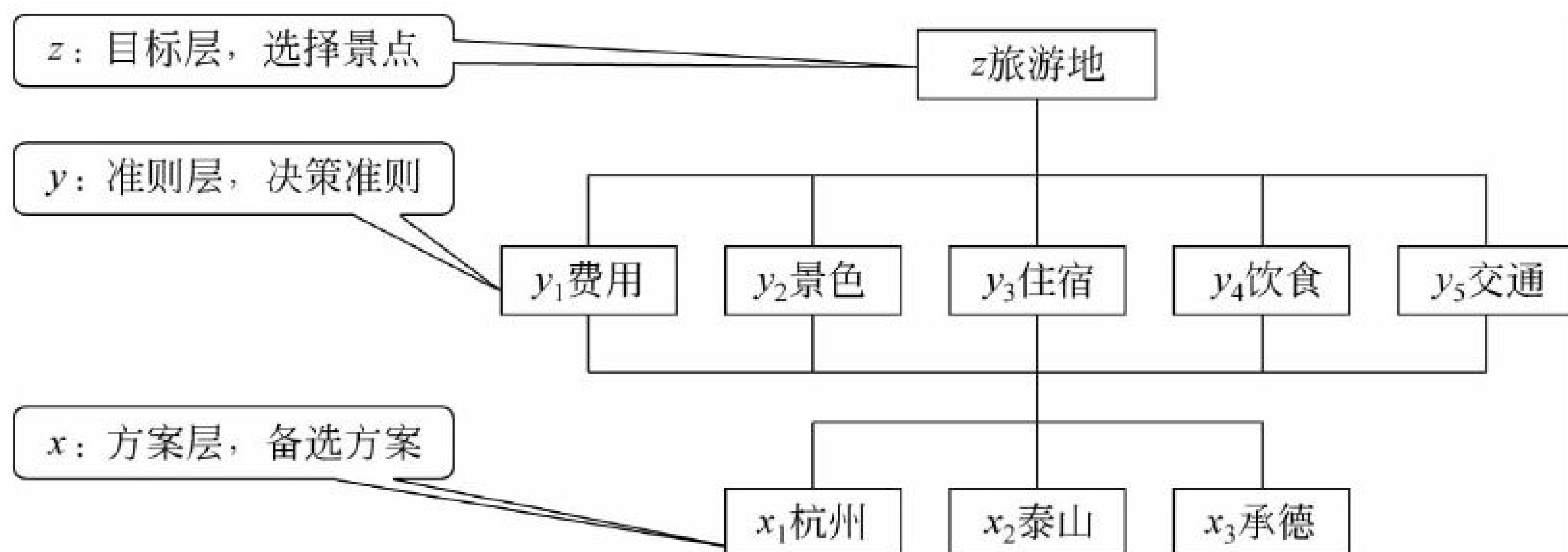


图 1-10 元素命名

下面构建两两比较的因素判断矩阵，以确定准则层对目标层的判断矩阵 $\boldsymbol{A}$ 为例。对于选择旅游地来说，若认为因素“y_1 费用”比“y_2 景色”稍稍重要，则矩阵 $\boldsymbol{A}$ 的第一行第二列元素为 2，第二行第一列元素为 1/2。若认为因素“y_1 费用”比“y_3 住宿”强烈重要，则矩阵 $\boldsymbol{A}$ 的第一行第二列元素为 7，第二行第一列元素为 1/7。其他类推，得判断矩阵 $\boldsymbol{A}$ 如下所示（判断矩阵因决策者不同而异）。

$$\boldsymbol{A}=\begin{bmatrix} 1 & 2 & 7 & 5 & 5 \\ 1/2 & 1 & 4 & 3 & 3 \\ 1/7 & 1/4 & 1 & 1/2 & 1/3 \\ 1/5 & 1/3 & 2 & 1 & 1 \\ 1/5 & 1/3 & 3 & 1 & 1 \end{bmatrix}$$

经计算可得，$\boldsymbol{A}$ 有最大特征值 $\lambda_{\max}=5.019$。$\lambda_{\max}$ 对应的归一化的特征向量为 $\vec{a}=(0.48,0.26,0.05,0.10,0.11)$。

计算得

$$CI = \frac{\lambda - 5}{5 - 1} = \frac{0.019}{4} = 0.00475$$

$$CR = 0.00475/1.12 = 0.004246 < 0.1$$

因此，$\boldsymbol{A}$ 有满意的一致性，$\vec{a}$ 即为准则层对目标层的单排序权重。

同理可得备选方案层对各决策准则的判断矩阵，检验一致性、计算排序权重如下所示。

$$\boldsymbol{B}_1 = \begin{bmatrix} 1 & 1/5 & 1/8 \\ 5 & 1 & 1/3 \\ 8 & 3 & 1 \end{bmatrix}, \quad \lambda_1 = 3.004, \quad \vec{b}_1 = \begin{bmatrix} 0.067 \\ 0.272 \\ 0.661 \end{bmatrix}$$

$$\boldsymbol{B}_2 = \begin{bmatrix} 1 & 2 & 5 \\ 1/5 & 1 & 2 \\ 1/5 & 1/2 & 1 \end{bmatrix}, \quad \lambda_1 = 3.005, \quad \vec{b}_2 = \begin{bmatrix} 0.595 \\ 0.277 \\ 0.129 \end{bmatrix}$$

$$\boldsymbol{B}_3 = \begin{bmatrix} 1 & 1 & 3 \\ 1/3 & 1 & 3 \\ 1/3 & 1/3 & 1 \end{bmatrix}, \quad \lambda_1 = 3, \quad \vec{b}_3 = \begin{bmatrix} 0.429 \\ 0.429 \\ 0.142 \end{bmatrix}$$

$$\boldsymbol{B}_4 = \begin{bmatrix} 1 & 3 & 4 \\ 1/3 & 1 & 1 \\ 1/4 & 1 & 1 \end{bmatrix}, \quad \lambda_1 = 3.009, \quad \vec{b}_4 = \begin{bmatrix} 0.633 \\ 0.193 \\ 0.175 \end{bmatrix}$$

$$\boldsymbol{B}_5 = \begin{bmatrix} 1 & 1 & 1/4 \\ 1 & 1 & 1/4 \\ 4 & 4 & 1 \end{bmatrix}, \quad \lambda_1 = 3, \quad \vec{b}_5 = \begin{bmatrix} 0.166 \\ 0.166 \\ 0.668 \end{bmatrix}$$

即准则 y 对目标 Z 有判断矩阵 $\boldsymbol{A}$，排序权重 $\vec{a}$。方案 x 对准则 y_j 有判断矩阵 $\boldsymbol{B}_j$，排序权重 $b_j=(b_{1j},b_{2j},b_{3j})'$。

记 $\boldsymbol{B}=(b_1,b_2,\cdots,b_5)$，$CI_j(x)$：$x$ 对 y_j 的 CI；$RI_j(x)$：x 对 y_j 的 RI；$CI_Z(x)$：x 对 Z 的 CI；$RI_Z(x)$：x 对 Z 的 RI。则排序的总一致性检验如下：

$$CI_Z(\vec{x}) = \sum_{j=1}^{5} a_j CI_j(\vec{x})$$

方案对目标的总排序求解方法如下所示。

$$\vec{w} = \sum_{j=1}^{5} a_j \vec{b}_j = \boldsymbol{B}\vec{a}$$

代入数据，求得 $\vec{w}=(0.293, 0.311, 0.446)'$。其中，$\vec{w}_3=0.446$ 为最大，即根据层次总排序权值，最满意的旅游地为承德。

为方便理解，整理层次总排序如表 1-5 所示。

表 1-5　层次总排序

准则		费用	景色	住宿	饮食	交通	总排序权值
准则层权值		0.48	0.26	0.05	0.10	0.11	
方案层	杭州	0.067	0.595	0.429	0.633	0.166	0.293
单排序	泰山	0.272	0.277	0.429	0.193	0.166	0.311
权值	承德	0.661	0.129	0.142	0.175	0.668	0.446

利用 MATLAB 软件编写计算程序如下所示。

```
clc
RI=[0  0  0.58  0.90  1.12  1.24  1.32  1.41  1.45];
B1=[1  1/5  1/8;  5  1  1/3;  8  3  1];
n1=size(B1);
n=n1(1);%n=rank(A);
[V,D]=eig(B1);
eigenvalue=diag(D);lamda=eigenvalue(1);%lamda=D(1,1);
CI=(lamda-n)/(n-1);
CR=CI/RI(n);
b1=V(:,1)/sum(V(:,1))
%类似可以计算 a,b2,...,bn
%下面计算总排序
%B=[b1 b2 b3];
%w_sum=B* a;
%w_sum_final=w_sum/sum(w_sum);
```

学习信息库 2——层次分析法的优缺点

在应用层次分析法研究问题时,遇到的主要困难有两个:①如何根据实际情况抽象出较为贴切的层次结构;②如何将某些定性的量做比较接近实际的定量化处理。

层次分析法对人们的思维过程进行了加工整理,提出了一套系统分析问题的方法,为科学管理和决策提供了较有说服力的依据。但层次分析法也有其局限性,主要表现在:第一,它在很大程度上依赖于人们的经验,主观因素的影响很大,它至多只能排除思维过程中的严重非一致性,无法排除决策者个人可能存在的严重片面性;第二,比较、判断过程较为粗糙,不能用于精度要求较高的决策问题。AHP 至多只能算是一种半定量(或定性与定量结合)的方法。

1. 层次分析法的优点

(1) 系统性的分析方法

层次分析法把研究对象作为一个系统,按照分解、比较判断、综合的思维方式进行决策,成为继机理分析、统计分析之后发展起来的系统分析的重要工具。系统的思想在于不割断各个因素对结果的影响,而层次分析法中每一层的权重设置最后都会直接或间接影响到结果,而且在每个层次中的每个因素对结果的影响程度都是量化的,非常清晰、明确。这种方法尤其可用于对无结构特性的系统评价,以及多目标、多准则、多时期等的系统评价。

(2) 简洁、实用的决策方法

简洁、实用的决策方法既不单纯追求高深数学,又不片面地注重行为、逻辑、推理,而是把定性方法与定量方法有机地结合起来,使复杂的系统分解,能将人们的思维过程数学化、系统化,便于人们接受,且能把多目标、多准则,又难以全部量化处理的决策问题化为多层次单目标问题,通过两两比较,确定同一层次元素相对上一层次元素的数量关系后,进行简单的数学运算。即使是具有中等文化程度的人,也可了解层次分析的基本原理和掌握它的基本步骤,计算也非常简便,并且所得结果简单、明确,容易被决策者了解和掌握。

（3）所需定量数据信息较少

层次分析法主要是从评价者对评价问题的本质、要素的理解出发，比一般的定量方法更讲求定性的分析和判断。由于层次分析法是一种模拟人们决策过程的思维方式的一种方法，层次分析法把判断各要素的相对重要性的步骤留给了大脑，只保留人脑对要素的印象，化为简单的权重进行计算。这种思想能处理许多用传统的最优化技术无法着手的实际问题。

2. 层次分析法的缺点

（1）不能为决策提供新方案

层次分析法的作用是从备选方案中选择较优者。这个作用正好说明了层次分析法只能从原有方案中选取，而不能为决策者提供解决问题的新方案。这样，在应用层次分析法的时候，可能会有一个情况，就是我们自身的创造能力不够，造成我们尽管在想出来的众多方案里选了一个最好的，但其效果仍然不如企业做出来的好。而对于大部分决策者来说，如果一种分析工具能代替其分析出在已知方案里的最优者，指出已知方案的不足，甚至提出改进方案，这样的分析工具才是比较完美的。显然，层次分析法没能做到这一点。

（2）定量数据较少，定性成分多，不易令人信服

在如今对科学的方法的评价中，一般都认为一门科学需要比较严格的数学论证和完善的定量方法。但现实世界的问题和人脑考虑问题的过程在很多时候并不是能简单地用数字来说明一切的。层次分析法是一种带有模拟人脑的决策方式的方法，必然带有较多的定性色彩。这样，当一个人应用层次分析法来做决策时，其他人就会说：为什么会是这样？能不能用数学方法来解释？如果不可以的话，你凭什么认为你的这个结果是对的？你说你在这个问题上认识比较深，我也认为我的认识比较深，可我和你的意见是不一致的，以我的观点做出来的结果也和你的不一致。这个时候，该如何解决？

比如，对于一件衣服，我认为评价的指标是舒适度、耐用度。这样的指标，对于女士们来说，估计是比较难接受的，因为女士们对衣服的评价一般认为美观度是最主要的，对耐用度的要求比较低，甚至可以忽略不计。因为一件便宜又好看的衣服，穿一次也值了，根本不考虑它是否耐穿。这样，对于一个我原本分析的“购买衣服时的选择方法”的题目，充其量也就只是“男士购买衣服的选择方法”。也就是说，定性成分较多的时候，可能这个研究最后能解决的问题比较少。

对于上述问题，其实是有办法解决的。如果说我的评价指标太少了，把美观度加进去，就能解决比较多问题了。指标还不够？再加嘛！还不够？再加！还不够?！不会吧？分析一个问题的时候考虑那么多指标，不觉得辛苦吗？大家都知道，对于一个问题，指标太多了，反而更难确定方案。这就引出了层次分析法的第二个不足之处。

（3）指标过多时，数据统计量大，且权重难以确定

若希望解决较普遍的问题，指标的选取数量可能随之增加。这就像利用系统结构理论分析一般系统的结构，要搞清楚关系环，就要分析到基层次；而要分析到基层次上的相互关系，需要确定的关系就非常多。指标的增加意味着要构造层次更深、数量更多、规模更庞大的判断矩阵。那么，就需要对许多指标进行两两比较。由于一般情况下层次分析法的两两比较是用1～9来说明其相对重要性，如果有越来越多的指标，对每两个指标之间的重要程度的判断可能出现困难，甚至会对层次单排序和总排序的一致性产生影响，使一致性检验不能通过。也就是说，由于客观事物的复杂性，或对事物认识的片面性，通过所构造的判断矩

阵求出的特征向量(权值)不一定是合理的。不能通过,就需要调整,在指标数量多的时候,这是个很痛苦的过程,因为根据人的思维定式,会觉得这个指标应该比那个重要,于是比较难调整过来,也不容易发现指标的相对重要性的取值里到底是哪个有问题,哪个没问题。可能花了很多时间,仍然不能通过一致性检验;更糟糕的是,根本不知道哪里出现了问题。也就是说,层次分析法里面没有办法指出在判断矩阵里是哪个元素出了问题。

学习信息库 3——优秀论文

学生宿舍设计方案的分析与评价

(2010 高教社杯全国大学生数学建模竞赛 D 题)

摘要　本文运用层次分析法建立模型,并利用 MATLAB 软件求解计算。本问题的难点在于因素判断矩阵的建立和影响因素的量化。因为不同的影响因素在建筑行业上的重要性不同,不能凭主观判断给出因素判断矩阵,而是应符合建筑设计行业的专业规范,所以本文采用专家调查法设计了宿舍设计影响因素调查问卷,邀请两位建筑行业的资深专家完成此份调查问卷,得到因素判断矩阵和影响因素量化的条件。在此基础上扬长避短,给出更加合理的评价方法。本文运用层次分析法,给出影响因素权重的总排序,并进行一致性检验,最终给出四种设计方案的得分总排序,确定出最优秀的设计方案是方案四。同时,本文利用软件图形用户界面(GUI)设计了一个小软件,该软件操作简单、使用方便,具有一定的实用性和一般性。

关键词　专家调查法、调查问卷、层次分析法、MATLAB 软件、图形用户界面 GUI

一、问题重述(略)

参见学习任务页-zh2。

二、模型假设及符号说明

2.1　模型假设

(1) 根据题目信息和专家调查问卷的结果,采用题中给出的 10 个影响因素。
(2) 假设使用面积只计算套间内面积。
(3) 假设四个宿舍设计方案的建筑结构形式是相同的。

2.2　符号说明

CI、RI、CR:分别表示一致性指标、随机一致性指标和一致性检验指标。

Z:建立层次结构模型的决策目标,即评价学生宿舍设计方案。

S_i:建立层次结构模型的因素层,S=(经济性 舒适性 安全性)($i=1,2,3$)。

M_j:建立层次结构模型的子因素层,M=(建设成本　运行成本　收费标准　人均面积　使用方便　互不干扰　采光　通风　人员疏散　防盗)($j=1,2,\cdots,10$)。

N_i:第 i 个设计方案($i=1,2,3,4$)。

Z-S:经济性、舒适性和安全性三方面影响因素层 S 对目标 Z 的因素判断矩阵。

S_i-M：S_i 层子因素对 S_i 层的因素判断矩阵（$i=1,2,3$）。

w_0：与矩阵 Z-S 最大特征根对应的特征向量归一化后取值。

w_i：与矩阵 S_i-M 最大特征根对应的特征向量归一化后取值（$i=1,2,3$）。

C_j：四种设计方案对第 j 个影响因素的权重规范化后得分向量（$j=1,2,\cdots,10$）。

H、R、V、F：分别表示四种设计方案中的建筑面积、学生总人数、寝室数量和一间寝室内的人数。

三、模型建立与分析求解

3.1　问题的分析——论文整体方法及思路分析

3.1.1　因素判断矩阵的建立——宿舍设计影响因素调查问卷的设计

本文设计了宿舍设计影响因素调查问卷，邀请两位建筑行业的资深专家完成此份调查问卷。请专家给出最符合建筑设计规则的影响因素权重（详见 3.2.1 小节）和各影响因素在建筑设计专业上如何量化（详见 3.2.2 小节）。

3.1.2　模型的扩展——MATLAB 软件图形用户界面 GUI

通过分析可以得出，在实际设计过程中，应根据不同学校、不同住宿人群等情况综合考虑影响因素的相对重要性，只考虑固定的权重是不合理的。所以，本文设计了 MATLAB 软件图形用户界面 GUI，只需根据实际情况在软件界面里输入影响因素的矩阵，即可得出权重；再单击“比较结果”按钮，即可得出设计方案的总排序（详见 3.2.3 小节）。

3.2　模型的建立与结果分析

本文主要从宿舍楼的经济性、舒适性和安全性三个方面，综合评价和比较四种典型的学生宿舍设计方案，运用层次分析法建立模型。因此，本文将建立层次分析模型。

3.2.1　影响因素权重的计算——层次分析法

1. 层次结构图的建立

1）目标层的确定

本文主要对四个设计方案进行量化评价和比较，即评价宿舍设计方案为目标层。

2）因素层的确定

通过对本题的分析可知，需从经济性、舒适性和安全性三方面评价设计方案，所以将经济性、舒适性和安全性作为第一层影响因素；还要考虑三个影响因素下三组子因素（共 10 个影响因素）的影响权重。

综上所述，给出层次结构图，如图 1 所示。

2. 因素判断矩阵的建立及一致性检验

1）因素判断矩阵建立的方法——宿舍设计影响因素调查问卷

（1）调查原因及目的

因为本题涉及建筑学专业领域，不同的影响因素在具体的建筑设计案例上的权重不同，每个影响因素量化的衡量条件也要符合建筑设计专业的规则，所以本文设计了宿舍设计影响因素调查问卷，邀请两位建筑行业的资深专家完成此份调查问卷，请专家给出最专业和最

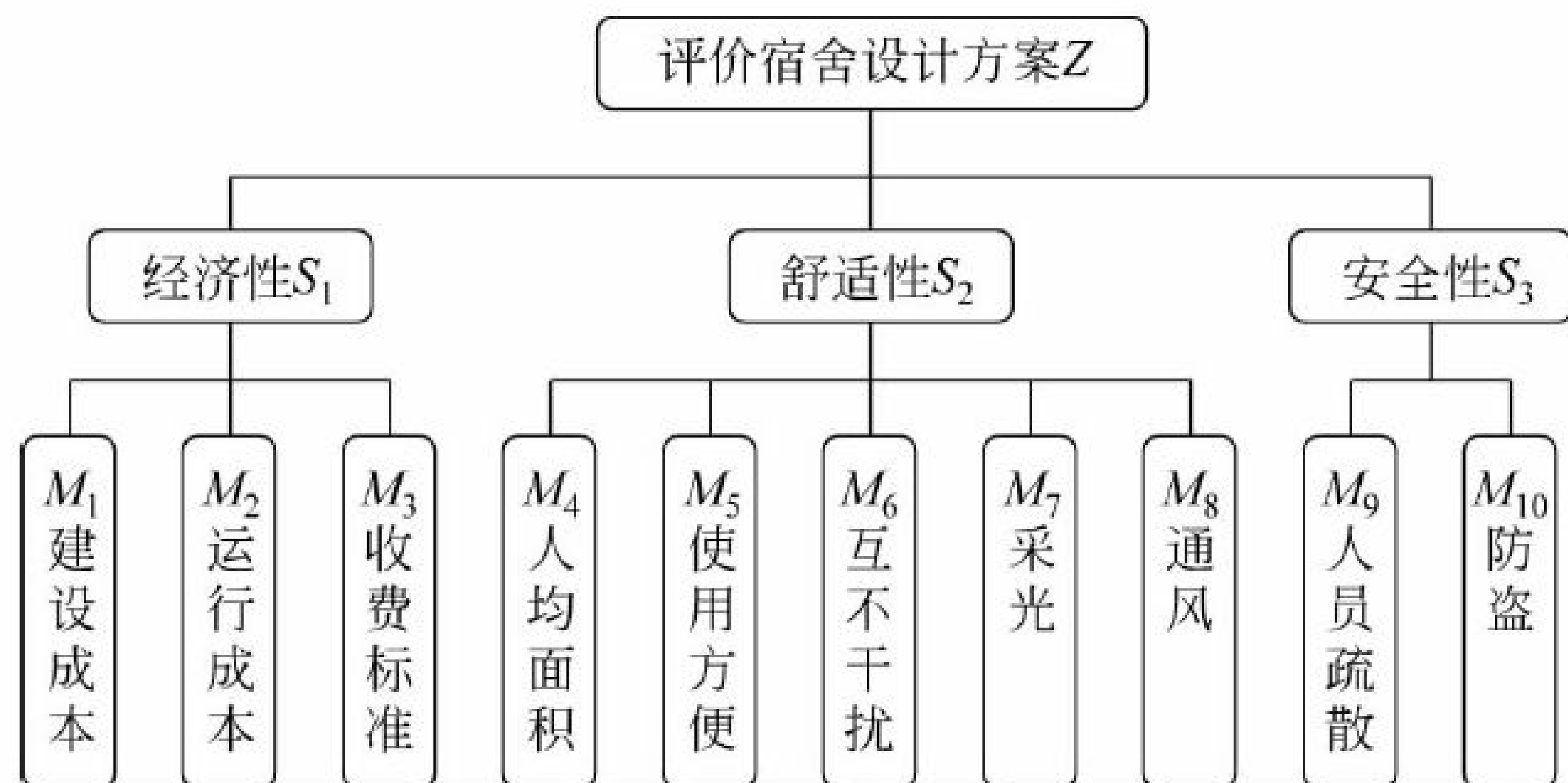


图 1　层次结构图

符合建筑设计规则的影响因素权重和各影响因素在建筑设计专业上的衡量条件。

通过调查，我们希望得出各影响因素的权重和各影响因素在建筑设计专业上的衡量条件，在与专家的沟通过程中得到更多参考资料。

（2）调查结果

通过调查，一共收回两份有效问卷，但是调查问卷一给出的因素判断矩阵的一致性检验未通过，所以本文采用调查问卷二给出的因素判断矩阵。

（3）调查问卷设计的优点与不足

通过调查问卷的形式，可以和有丰富工作经验的专家直接沟通，确定层次分析法中各影响因素的权重，以及各影响因素如何量化等关键资料。

另外，通过调查，了解专家之前进行建设方案评价时采用的方法，本文在评价设计方案时可以吸取这些方法的优点，完善自己评价设计方案时的方法。

2）确定层次因素的单层排序

层次单排序计算问题可以归结为计算判断矩阵的最大特征根及特征向量的问题。

综合以上分析、求解，得出判断矩阵 Z-S 和判断矩阵 S-M 的最大特征根所对应的归一化特征向量，再综合各类数据，得到各影响因素的单层权重，如图 2 所示。

由图 2 可以看出，对目标来说，在经济性、舒适性和安全性三方面影响因素中，经济性的权重最大，而舒适性和安全性两项影响因素所占的权重较小，即重要性程度排序为：经济性＞舒适性＞安全性。

对经济性来说，建设成本、运行成本和收费标准三个影响因素中，收费标准的权重占70％多的比例，建设成本和运行成本占很小的权重，即重要性程度排序为：收费标准＞建设成本＞运行标准。

同理可得，对于舒适性和安全性来说，重要性程度排序分别为：使用方便＞人均面积＞互不干扰＞通风＞采光；人员疏散＞防盗。

3）总排序及其一致性检验

层次总排序就是利用层次单排序的结果计算出各层次的组合权值，即每一个判断矩阵的各因素针对目标层（最上层）的相对权重。层次总排序的一致性检验之后，得到层次总排序。

求得第二层 10 个影响因素对于总目标 Z 的权重为

$$w_4=(0.1122\quad 0.0618\quad 0.4706\quad 0.0830\quad 0.1454\quad 0.0448\quad 0.0111\quad 0.0247\quad 0.0821\quad 0.0274)'$$

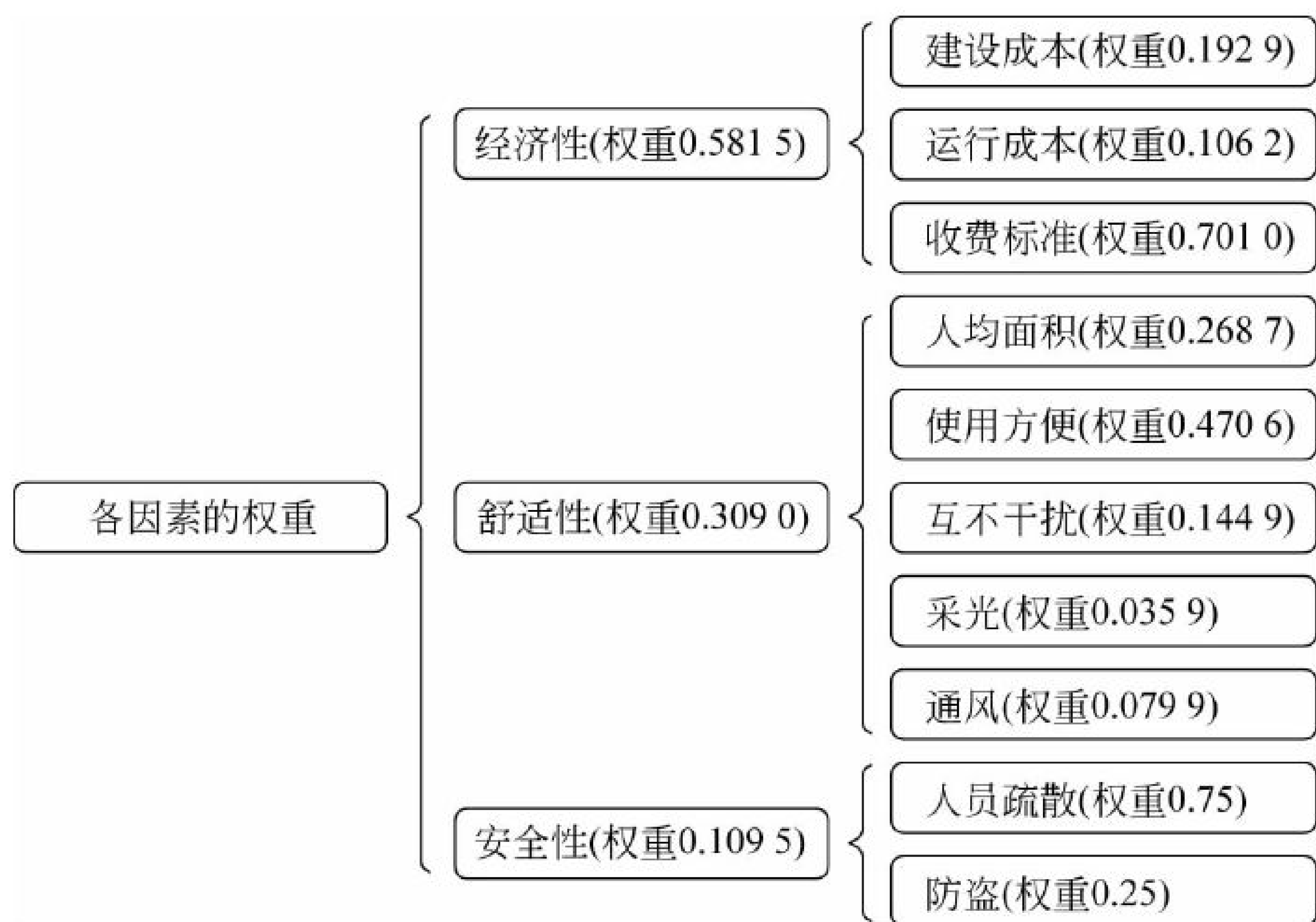


图 2　各影响因素的单层权重

即在宿舍设计方案中，10 个影响因素的重要性排序为：①收费标准；②使用方便；③建设成本；④人均面积；⑤人员疏散；⑥运行成本；⑦互不干扰；⑧防盗；⑨通风；⑩采光。

因为三方面影响因素 $S_i(i=1,2,3)$ 的排序已完成，所以得到权重值 w_0，与 w_0 对应的本层次因素 $M_j(j=1,2,\cdots,10)$ 单排序结果为 $w_i(i=1,2,3)$，总排序为 w_4。

为评价层次总排序计算结果的一致性，用下述公式进行检验。

$$CI_{总}=\sum_{i=1}^{3}w_iCI_i,\quad RI_{总}=\sum_{i=1}^{3}w_iRI_i,\quad CR_{总}=\frac{CI_{总}}{RI_{总}}$$

经一致性检验，指标 $CR_{总}<0.1$，说明有满意的一致性。

3.2.2　影响因素的量化

1. 正、逆指标处理方法

影响因素指标对评价对象的作用分为正影响及逆影响两种。若将这两类指标混为一谈，分析出的综合指标就没有任何意义。所以，需要将影响因素指标对评价对象的作用统一。常用的统一正、逆指标的方法有最大值做差法和倒数法两种。

若有 n 个评价对象，记 X_i 是第 i 个评价对象的属性数值。设 Y_i 是运用最大值做差法处理后 X_i 的属性值$(i=1,2,\cdots,n)$，则有

$$Y_i=\max_i\{X_i\}-X_i\quad(i=1,2,\cdots,n)$$

设 Z_i 是运用倒数法处理后 X_i 的属性值$(i=1,2,\cdots,n)$，则有

$$Z_i=1/X_i\quad(i=1,2,\cdots,n)$$

2. 数据规范化方法

属性 X_i 之间由于量纲及量级不同而存在不公平性，因此，需要对指标进行无量纲化处理。指标的无量纲化叫作指标数据的标准化、规范化。

若有 n 个评价对象，记 X_i 是第 i 个评价对象的属性数值。设 U_i 是属性 X_i 的规范化属性值$(i=1,2,\cdots,n)$，则有

$$U_i=\frac{X_i-\min_i\{X_i\}}{\max_i\{X_i\}-\min_i\{X_i\}}$$

3. 10 个影响因素的量化

1）建设成本

综合调查问卷信息，可得建设成本与建设面积和建筑结构形式两方面因素有关，但是由于四个方案设计图纸中都没有给出具体的建筑结构形式，所以本文假设它们的建筑结构形式是相同的，故而建筑成本只与建设面积有关。因为建设成本是建设宿舍楼所投入的费用，而建设面积越大，所花费的建设成本越高，故假设建设成本与建筑面积成正比，即 $M_1 \propto H$。其中，M_1 为建设成本，H 为建筑面积。建设成本越高，获得的利润越少。将 H 中的数据进行正、逆指标处理，采用取倒数的方法（详见 3.2.2 小节（一）），再进行规范化（详见 3.2.2 小节（二）），得出四种设计方案在建设成本中的得分向量，即

$$C_1 = (1 \quad 0 \quad 0.09 \quad 0.21)$$

2）运行成本

综合调查问卷信息可得，在建筑学方面，运行成本与水费、电费、排暖费、管理费、维修费五项因素有关。

水费取决于学生的用水量，即与学生人数 R 有关。学生人数越多，用水量越大，故假设水费与学生人数成正比，即 $A_1 \propto R$。其中，A_1 为水费，R 为学生人数，k 表示常量系数，$i=1,2,\cdots,4$。进行规范化，得出四种设计方案在水费方面的得分向量为

$$T_1 = (0.54 \quad 0.92 \quad 1 \quad 0)$$

电费取决于每个房间的用电量，与房间数量 V 有关。房间的数量越多，电费越多，故假设电费与房间的数量成正比，即 $A_2 \propto V$。其中，A_2 为电费，V 为房间数量。进行规范化，得出四种学生宿舍的设计方案在电费方面的得分向量为

$$T_2 = (0.03 \quad 1 \quad 0.48 \quad 0)$$

排暖是指整个宿舍楼的供暖。建筑面积越大，供暖费越多，故假设排暖费与建筑面积成正比关系，即 $A_3 \propto H$。其中，A_3 为排暖费，H 为建筑面积。进行规范化，得出四种设计方案在排暖费中的得分向量为

$$T_3 = (0 \quad 1 \quad 0.76 \quad 0.57)$$

管理费是指宿舍管理需要的费用，故与建筑面积有关。建筑面积越大，管理费用越高，故假设管理费与建筑面积成正比，即 $A_4 \propto H$。其中，A_4 为排暖费，H 为建筑面积。进行规范化，得出四种学生宿舍的设计方案在管理费中的得分向量为

$$T_4 = (0 \quad 1 \quad 0.76 \quad 0.57)$$

维修费用于对宿舍的设施进行维修、管理，宿舍所住的学生人数越多，则维修管理费用越高，故假设维修费正比于学生人数，即 $A_5 \propto R$。其中，A_5 为维修费，R 为学生人数。

运行成本是指宿舍楼投入使用后，学校需要支付的各种费用，即五项因素的得分向量之和，即

$$M_2 = \sum_{i=1}^{5} A_i = k_1 H + k_2 V + k_3 R$$

运行成本应越少越好。将 M_2 中的数据进行正、逆指标互换处理，采用取倒数方法，再进行规范化，得出四种学生宿舍的设计方案在运行成本中的得分向量为

$$C_2 = (0.46 \quad 0 \quad 0.06 \quad 1)$$

3）收费标准

根据在北京教育委员会网站中查询的公办高校住宿费标准，分别给出四种学生宿舍的

设计方案中每间宿舍的收费标准，如表1所示。

表1 四种学生宿舍的收费标准

宿舍	宿舍一	宿舍二	宿舍三	宿舍四
收费标准/(元/年)	900	1 500	1 200	1 800

再进行规范化，得出四种学生宿舍的设计方案在收费标准中的得分向量为

$$C_3=(0\quad 0.67\quad 0.33\quad 1)$$

4）人均面积

从舒适性的角度看，人均面积是指在宿舍中每人的平均使用面积，即使用面积与学生人数的比例。因此，对于学生人数少、使用面积大的宿舍，人均面积的权重肯定大。假设使用面积只计算套间内的面积，即 $M_4 \propto D/R$。其中，M_4 为人均面积，D 为使用面积，R 为学生人数。进行规范化，得出四种学生宿舍的设计方案在人均面积中的得分向量为

$$C_4=(0\quad 0.79\quad 0.56\quad 1)$$

5）使用方便

使用方便在宿舍环境中主要考虑卫生间及盥洗室，要考虑到卫生间里人均共用马桶个数，以及盥洗室里人均盥洗槽的使用个数。使用方便是人均共用马桶个数与人均公用盥洗槽个数因素的得分向量之和，即 $M_5 \propto E/R+O/R$。其中，M_5 为使用方便，E 为马桶个数，O 为盥洗槽个数，R 为学生人数。

将结果进行规范化，得出四种设计方案在使用方便中的得分向量为

$$C_5=(0\quad 0.99\quad 0.88\quad 1)$$

6）互不干扰

互不干扰在宿舍环境中主要是受同一空间内人数的影响。在四种设计方案中，一间寝室的人数为总人数与房间数量的比值，即 $M_6 \propto R/V=F$。其中，M_6 为互不干扰，R 为学生人数，V 为房间数量。因为同一空间内的人数越少，互相干扰就越少，将 F 中的数据进行正、逆指标互换处理，采用取倒数的方法，再进行规范化，得出四种宿舍设计方案在互不干扰中的得分向量为

$$C_6=(0\quad 0.33\quad 0.12\quad 1)$$

7）采光

通过综合分析两位专家给出的专业知识可以得出，宿舍的采光主要靠窗户的面积占整个空间面积的比例来衡量，因为四个宿舍设计方案中都没有明确给出窗户的具体高度，所以假设四个宿舍设计方案中的窗户高度统一为 h；因为每间寝室的布局都是一样的，所以四个设计方案中采光的量化只需考虑一间寝室，即只需得出每个设计方案中一间寝室的窗户面积与整个寝室空间面积的比例即可。

假设采光与窗户的面积占寝室面积的比例成正比，即 $M_7 \propto Bh/P$。其中，M_7 为采光，B 为窗户的宽度，h 为窗户的高度，P 为寝室面积。规范化后的向量为

$$C_7=(1\quad 1\quad 1\quad 0)$$

8）通风

通过查找建筑设计方面的规则可知，建筑物室内应有与室外空气直接流通的窗户或开口；否则应设置有效的自然通风道或机械通风设施。采用直接自然通风应符合下列规定：自然通风道应设于窗户或与进风口相对的一面，故假设通风与窗户面积成正比，即 $M_8 \propto B_i h$。其中，M_8 为通风，B_i 为窗户的宽度，h 为窗户的高度。规范化后的向量为

$$C_8=(0\quad 0\quad 1\quad 0.22)$$

9）人员疏散

通过调查问卷分析得出，人员疏散与两个影响因素有关：人均楼梯宽度，以及离楼梯最远距离的寝室到楼梯的距离。楼梯宽度等于楼梯个数与楼梯宽度的乘积，人均楼梯宽度等于楼梯宽度占总学生人数的比例，可得出四个宿舍设计方案的人均楼梯宽度 $K=(0.01\quad 0.04\quad 0.04\quad 0.06)$，规范化后的向量为

$$P=(0\quad 0.6\quad 0.6\quad 1)$$

在发生紧急情况而需要人员安全疏散时，关键在于最后一个逃生的人所用的时间，以确保在最短时间内所有人都能逃生。最后一个逃生的人所走的路线就是离楼梯最远距离的寝室到楼梯的距离。由此计算出四个方案中离楼梯最远距离的寝室到楼梯的距离 $L=(13.6\quad 28.8\quad 27.9\quad 29.3)$，$L$ 的值越小越好，说明能在更短的距离和时间里逃生。所以，运用正逆指标得出 $L=(0.07\quad 0.03\quad 0.04\quad 0.03)$，规范化后的向量为 $Y=(1\quad 0\quad 0.33\quad 0)$。综合上述两个影响因素，得出人员疏散的量化公式为 $M_9\propto I_iK_i/R+1/L$。其中，M_9 为人员疏散，I 为楼梯个数，K 为楼梯宽度，R 为学生人数，L 为离楼梯最远距离的寝室到楼梯的距离。将影响人员疏散的两个因素规范化后的向量相加，得出四个方案中人员疏散的得分，再进行规范化，得出最后得分为

$$C_9=(1\quad 0\quad 0.825\quad 1)$$

10）防盗

防盗的影响因素有首层护栏的安装、每层出入口的个数及位置、监视探头及监视器的安装三方面，但是由于四种宿舍设计方案中没有给出三方面具体的配置情况，故可以假设各方案的防盗措施都是一样的，因此各方案在防盗方面的权重都是相同的，可以不考虑在内。不妨设四种宿舍设计方案在防盗方面的量化得分为

$$C_{10}=(0\quad 0\quad 0\quad 0)$$

3.2.3 设计方案的总排序

层次总排序就是利用层次单排序的结果计算出各层次的组合权值，即每一个判断矩阵各因素针对目标层（最上层）的相对权重。将四个宿舍设计方案对 10 个影响因素的权重向量$(C_1,C_2,\cdots,C_{10})$建立矩阵 $\boldsymbol{C}$，故四个宿舍设计方案的最终得分分别等于其在 10 个影响因素上的得分判断矩阵 $\boldsymbol{U}$ 与 10 个影响因素的权重向量 w_4 分别相乘的和。

最后得出四个宿舍设计方案的最终得分：方案一的得分为 0.233 8，方案二的得分为 0.451 8，方案三和方案四的得分分别为 0.427 7 和 0.853 7。

可以看出，通过对影响宿舍设计方案的因素进行量化和综合分析，评价出最优秀的是方案四。

3.3 模型改进——引入 MATLAB 软件图形用户界面 GUI

在实际设计过程中，要根据不同学校、不同住宿人群等情况，综合考虑影响因素的相对重要性，只考虑本文中专家给出的固定权重是不合理的，故本文引入 MATLAB 软件图形用户界面 GUI。

本文运用图形用户界面 GUI 编程建立一个半自动的复杂系统决策模型。半自动是指使用者只需根据实际情况手动输入各因素的矩阵并单击按钮，MATLAB 程序运行相应语句，自动得出结果。

3.4 模型的推广——MATLAB软件程序

MATLAB软件的操作十分便捷，步骤如下：只需根据实际情况在软件界面框Ⅰ里相应位置输入影响因素的判断矩阵，单击框Ⅱ的“计算权重”按钮，即可在框Ⅲ中得出权重；再单击框Ⅳ中的“计算比较结果”按钮，可得出框Ⅴ中的各设计方案总得分。得分最高的，即为最优方案。

教学方法和教学策略

教学方法	项目教学法		
师生角色	以学生为中心；教师是启发者和辅导者		
实施重点	任务驱动，让学生“在干中学”		
实施步骤	教学过程与内容	教学方法与策略	学时
1. 资讯	(1) 分析并明确工作、学习任务 (2) 引导学生搜集相关信息	引导文教学法、讲授法	1
2. 计划	(1) 了解工作计划要素 (2) 制订初步的工作计划	小组讨论法 问题引导法 小组总结	1
3. 决策	学生讨论、修改，确定工作计划	师生互动 共同探讨	1
4. 实施	(1) 完成因素指标层次模型 (2) 完成因素层权重计算 (3) 进行各因素指标量化计算 (4) 综合计算总得分，给出评价结果	实地观察法 头脑风暴法 思维导图 小组讨论法	6
5. 检查	(1) 对照工作计划及任务要求，检查综合评价方案的过程与结果 (2) 发现存在的问题与不足，并做出相应的修改	问题引导法 对比分析法	1
6. 评价	(1) 师生共同评价工作任务完成的效果和质量 (2) 针对不足提出改进措施与注意事项	小组讨论 展示	2

成果展示要求

展示成果	1. 物流中心选址模型(学习任务页-zh1)论文(Word排版) 2. 物流中心选址综合评价过程介绍(PPT)
展示方式	1. 每组把促销方案制作成幻灯片进行展示 2. 小组可以选出一个人进行展示，其他同学发现有不完善的地方，可以补充；也可以分工，每个同学展示方案的一部分
展示顺序	抽签决定
展示时间	每个小组的展示时间为7分钟，回答问题的时间为3分钟

续表

展示要求	物流中心选址模型论文	• 格式正确，排版规范 • 结构完整，有较强的逻辑性和系统性 • 重点突出，包含重要信息和关联点 • 计算准确，论述周详 • 假设合理，推导严密，结论理由充分 • 条理清楚；行文流畅
	PPT 展示	• 格式和视觉效果较好 • 内容完整，条理清楚，逻辑严谨 • 内容、语言和媒体的选择相辅相成 • 体现团队精神 • 语言富有表现力，易于理解 • 有眼神交流，表情生动 • 语速、语调适中

学习评价——个人工作过程报告

学生信息：________________

请在下表中认为描述相符的数字位置画上"√"。

1	小组工作中，个人感觉小组氛围如何？						
	非常好	1	2	3	4	5	非常不好
2	小组成员对本人的接受度如何？						
	完全接受	1	2	3	4	5	完全排斥
3	本人的行为是否赢得了小组的理解？						
	非常强	1	2	3	4	5	根本没有
4	在项目中是否学到了东西？						
	非常多	1	2	3	4	5	根本没有
5	对小组工作结果是否满意？						
	非常满意	1	2	3	4	5	完全不满意
6	感觉老师对我们小组工作						
	非常有帮助	1	2	3	4	5	没有帮助和引导

通过下面的调查表，对小组中个人的工作进行评价。

在我的团队中，我能……	评　　价
专业能力 • 按照专业和实际情况展示学习和工作成果 • 独立或在小组中按照时间要求设计工作步骤 • 充分利用专业信息来源和工作辅助工具 • 想出新创意，给出创造性建议 • 区别重要信息和非重要信息	

续表

在我的团队中，我能……	评　　价
方法能力 • 获取、阅读、分析、理解信息资料 • 策划工作步骤，在规定时间内实施 • 根据目标，采用基本的工作技巧 • 组织新观点，对比不同意见和评价 • 采用适当的方法展示自己的观点、想法和结果	
社会能力 • 接受并遵守小组内达成一致的交流和对话准则 • 为自己的观点提出理由和依据 • 能够对不同意见做出适当的反应 • 接受小组分配的任务，并参与任务分配工作 • 正确面对小组内的冲突，并找出解决方法 • 向其他学生提供帮助，向其他学生寻求帮助	
自我认识能力 • 自己检查工作成果是否合理、是否正确 • 自己认识到当前工作的步骤及缺点 • 了解个人在小组中的位置和贡献 • 自己确立合适的工作和行动目标	

学习评价——小组工作评价

每个小组中应该确定一名小组长，作为此次工作的负责人。每名组员应该在各项任务中轮流担任小组长。小组长要注意，是否考虑到工作中的每一步，并且实施。当小组工作偏离或背离主题时，应该为小组找到正确的方向。

请小组长在下表右侧各列中，自上而下，将已经完成的工作点用“√”的形式表示出来。

学习任务编号		1	2	3	4	5	6	7	zh1	zh2
计划阶段	分配任务，弄清任务要求									
	确定工作步骤									
	估计所需时间，制定工作时间表									
实施阶段	顺利完成任务									
	互相帮助，并提供咨询									
	认真、仔细地进行和题目相关的工作									
	定时检查工作进度									
	准时准备好报告									
评价阶段	客观地评价工作结果									
	客观地分析合作情况									
	总结对下次小组工作有用的方法									

对自己和小组其他成员的评价：小组中的每个成员都应在小组工作结束后进行此评价。

是否注意到以下小组规则?	学习任务编号(写出评价分数,5=极好,0=极差)								
	1	2	3	4	5	6	7	zh1	zh2
小组顺利地开展工作									
小组成员对工作任务详细讨论,并确定目标									
小组成员互相帮助,并且互相关注									
小组成员听取组员意见并且思考									
小组中的所有成员以同样的工作强度参与工作									
每个组员都记录小组及个人的工作过程									
出现问题时,认真讨论并共同解决									
小组成员专心于主题工作,而没有进行其他活动									
小组讨论中没有出现人身攻击或者谩骂的现象									

第二部分

数据的预处理

学习情境——学生考试成绩分析

表 2-1 所示是某大专院校招收的一个班 46 名学生的高考成绩数据。根据数据，请分析该班学生语、数、外三门课成绩及总成绩。首先，请分析每门课程的一般水平，并比较三门课程的一般水平；其次，请筛选出三门课程成绩都及格的学生名单；最后，以 10 为组距，结合图表，展示每门课程各分数段的分布情况。

表 2-1　某大专院校一个班 46 名学生的高考成绩数据

姓名	总成绩	语　文	数　学	英　语	姓名	总成绩	语　文	数　学	英　语
王婷	415.00	107.00	69.00	48.00	杨婷	272.00	96.00	56.00	41.00
闫冬	404.00	105.00	75.00	45.00	蒙丹	268.00	94.00	101.00	73.00
王雪	384.00	91.00	102.00	48.00	田娇	257.00	95.00	35.00	43.00
张欧	383.00	97.00	59.00	82.00	乔瑞	248.00	93.00	92.00	63.00
王菁	375.00	98.00	74.00	49.00	李松	235.00	96.00	87.00	52.00
郭楠	374.00	95.00	85.00	57.00	王冬	233.00	90.00	45.00	37.00
刘强	369.00	99.00	83.00	42.00	夏林	232.00	82.00	85.00	65.00
邢文	357.00	105.00	69.00	61.00	孙荣	231.00	82.00	61.00	88.00
李铭	356.00	91.00	66.00	62.00	王月	230.00	93.00	69.00	68.00
赵静	353.00	91.00	78.00	48.00	心瑶	225.00	91.00	83.00	51.00
王瑞	347.00	105.00	30.00	76.00	晓林	224.00	95.00	88.00	41.00
丁骧	341.00	100.00	73.00	51.00	杨雨	224.00	98.00	90.00	36.00
杨新	337.00	86.00	94.00	42.00	张红	223.00	86.00	76.00	61.00
李璐	327.00	83.00	76.00	65.00	徐卉	222.00	89.00	69.00	64.00
孙灵	324.00	97.00	73.00	56.00	石玉	222.00	100.00	58.00	64.00
孙淼	324.00	86.00	73.00	71.00	雨彤	221.00	80.00	71.00	70.00
思思	307.00	93.00	74.00	55.00	雪莲	220.00	97.00	61.00	62.00
源源	299.00	91.00	59.00	60.00	刘畅	217.00	93.00	34.00	90.00
王芳	281.00	93.00	115.00	73.00	张扬	189.00	87.00	58.00	44.00
龙杨	276.00	85.00	51.00	61.00	朱丽	188.00	59.00	60.00	69.00

续表

姓名	总成绩	语　文	数　学	英　语	姓名	总成绩	语　文	数　学	英　语
刘霞	184.00	75.00	50.00	59.00	杨芳	163.00	85.00	40.00	38.00
王奇	182.00	97.00	43.00	42.00	王楠	157.00	78.00	22.00	57.00
魏薇	169.00	74.00	45.00	50.00	贺帝	154.00	66.00	57.00	31.00

能力矩阵——数据的预处理

学生活动	教师控制/教师组织			自我导向/自我组织		
教师角色	讲授、辅助咨询			教练		
学习步骤	能力水平					
	认　知		应　用		创　新	
	A1	A2	B1	B2	C1	C2
	识　记	理　解	应　用	分　析	创　造	评　价
1. 数据预处理的目的与步骤	能记住数据预处理的目的及基本步骤	能理解数据的概念及其分类，能理解数据收集的概念	能判断已有数据的类型			
2. 数据的审核、筛选、排序	能记住数据审核、筛选、排序的目的	能理解数据审核、筛选及排序的具体内容及方法	能对已给的数据进行审核能用 Excel 对数据进行筛选与排序	能够对排序与筛选的结果进行简单的文字解释与总结	能够用 Excel 对数据进行高级筛选	能够评价自动筛选与高级筛选的优缺点
3. 数据的分组或分类	能记住分组或分类的含义，以及数据分组的原则	理解有哪些数值型数据的分组方法，每种方法的基本概念及分组要点	能对已有数据分出合理的“组”	能把数据归类到各组		
4. 数据展示：制表	能记住数据展示的意义、频数、频率等概念	能理解制作频数分布表的意义与具体方法，以及数据透视表的概念	能利用 Excel 制作频数分布表	能根据所制作的频数分布表对数据进行合理解释		
5. 数据展示：绘图	能记住常用的数据展示图有哪些	能理解绘图的意义及每种图形适合展示的数据分布类型	能应用 Excel 绘制常用的数据展示图	能针对实际问题绘制出合适的展示图	能对所绘制的展示图进行合理解释	
6. 数据的统计描述	能记住常用的描述统计量有哪些及其概念	能理解常用描述统计量所表达的含义及计算方法	能应用 Excel 计算常用的描述统计量	能对所计算出的描述统计量的结果进行合理解释，并形成文字	能综合比较、分析以上各种结果，形成一份论文	

学习步骤计划书

姓　　名		班　　级	
学习日期		汇报日期	
学习任务			
学习目标			
(1) 我的目标是什么？ **注意：**每一个目标都应该以"我能……"开始，并将每一个目标编号。请注明通过哪一个学习任务达到该学习目标。 (2) 我想要学会什么？ **注意：**要给出在哪个时间点达到这个目标。			
学习途径			
(1) 我通过哪些途径获得该目标的相关信息？ (2) 我需要什么材料？ (3) 我可以通过哪些方式完成该目标？ (4) 达到此目标的过程中会出现什么障碍？ (5) 在完成此目标的每一天中，我要完成什么任务？			
学习证明			
(1) 通过何种展示方式证明我达到这个目标？如论文写作、PPT 展示、流程图绘制等。 (2) 怎样检测我是否达到这个目标？ (3) 请给出预计展示成果的时间。			
学习评价			
(1) 对于学习结果，我是否满意？ (2) 我通过什么方法评价自己的学习质量？ (3) 我的学习信息是否充足？我如何克服上面提到的障碍？ (4) 我还能做些什么，以完善学习成果？			

学习任务页

学习任务页-1　数据预处理的目的与步骤

学生信息__________　　任务开始日期__________　　任务分值　10 分　

1-A1-1　何为数据预处理？数据预处理的目的是什么？

1-A1-2　数据预处理的基本步骤有哪些？

1-A2-1　何为数据？数据分哪几种类型？

1-A2-2　何为数值型数据？何为属性型数据？

1-B1-1　请判断学习情境中数据表 2-1 中所列数据的类型。

1-B1-2　表 2-2 所示是某教研室教师基本信息表，请判断其数据类型。

表 2-2　某教研室教师基本信息表

姓　名	性别	政治面貌	学　　历	职　称
王军	男	党员	硕士研究生	讲师
张珊	女	党员	本科	讲师
刘天明	男	党员	博士研究生	教授
陈赫	男	群众	硕士研究生	副教授
李琳	女	群众	硕士研究生	副教授
吴桐	女	群众	本科	讲师
武铭	男	党员	硕士研究生	教授

1-B1-3　某工程建设项目需要招标，共有 4 家单位参与投标。4 家投标单位的各项评价指标汇总信息如表 2-3 所示。请判断哪些是数值型数据，哪些是属性型数据。

表 2-3　投标方案评价指标汇总表

单位	报价/万元	工期/天	质量实绩	施工方案	企业信誉	项目经理
甲	921	230	优	优	好	一级
乙	917	246	良	优	很好	二级
丙	912	268	优	良	好	二级
丁	904	255	中	良	一般	一级

学习任务页-2　数据的审核、筛选、排序

学生信息__________　　任务开始日期__________　　任务分值____15 分____

2-A1　数据审核、筛选、排序的目的是什么？

2-A2-1　数据审核的具体内容和方法有哪些？

2-A2-2　数据的筛选包括哪些内容？

2-A2-3　数据排序的具体方法有哪些？

2-B1-1　请利用 Excel 进行数据筛选。

(1) 表 2-4 所示的是 10 名学生 4 门课程的期末考试成绩数据。请筛选出微积分成绩低于等于 75 分的学生,语文成绩前 3 名的学生,4 门课程的成绩都高于 80 分的学生。

表 2-4　10 名学生 4 门课程的期末考试成绩数据

姓名	微积分	大学语文	英语	经济学
张强	95	88	84	89
王阳	91	75	96	93
张雨	58	80	65	79
雍华	88	89	87	68
李颖	75	70	80	82
宋军	86	72	56	73
袁蕊	45	58	78	91
陈晨	89	89	93	78
常英	90	92	85	90
苏珊	96	95	89	88

(2) 表 2-5 所示的是 6 名学生的期中考试成绩登记表,请将各门课程成绩不及格的学生名单筛选出来。

表 2-5　6 名学生的期中考试成绩登记表

序号	姓名	语文	数学	英语	政治	历史	地理	生物	总分
1	小吴	63	84	73	69	75	50	52	466
2	小曲	43	57	70	56	62	75	71	434
3	小苏	82	88	67	91	72	81	81	562
4	小许	79	57	81	76	58	51	68	470
5	小齐	68	73	72	78	69	30	89	479
6	小单	48	87	71	67	73	71	85	502

2-B1-2　针对“任务 2-B1-1”第(2)题所列数据(见表 2-5),对各门课程的成绩及总成绩进行降序排列,并简单总结排序结果。

2-B2-1　请对“任务 2-B1-1”中第(1)题的筛选结果进行文字总结。

2-B2-2　请对“任务 2-B1-1”中第(2)题的筛选结果进行文字总结。

2-C1　表 2-6 所示的是通辽市上半年福特汽车销售的部分数据。根据数据资料,完成以下任务。

(1) 筛选所有销售价格大于 16 万元的银砂黑颜色的汽车信息。

(2) 筛选所有销售价格小于 20 万元,且库龄天数大于 200 的汽车信息。

(3) 筛选所有 1 月份销售的炫耀银颜色的汽车信息。

(4) 筛选所有销售价格小于 15 万元,且库龄天数小于 100 的钛晶灰颜色的汽车信息。

(5) 筛选所有销售价格大于 10 万元,且库龄天数大于 100 的炫耀银颜色的嘉年华汽车信息。

表 2-6　通辽市上半年福特汽车销售的部分数据

<table>
<tr><th>销售月份</th><th>子品牌</th><th>车 型 配 置</th><th>车型代码</th><th>颜　色</th><th>库龄天数</th><th>销售价格/万元</th></tr>
<tr><td rowspan="6">1 月</td><td rowspan="3">嘉年华</td><td>三厢 1.3L 手动风尚型</td><td>85744</td><td>炫耀银</td><td>236</td><td>23.86</td></tr>
<tr><td>三厢 1.3L 手动风尚型</td><td>85744</td><td>炫耀银</td><td>242</td><td>22.58</td></tr>
<tr><td>三厢 1.5L 自动时尚型</td><td>85708</td><td>炫耀银</td><td>253</td><td>19.68</td></tr>
<tr><td rowspan="3">蒙迪欧致胜</td><td>GTDI 200 豪华型浅内饰</td><td>86532</td><td>银砂黑</td><td>183</td><td>20.47</td></tr>
<tr><td>三厢 2.3L 自动时尚型</td><td>86526</td><td>银砂黑</td><td>152</td><td>19.88</td></tr>
<tr><td>GTDI 200 时尚型</td><td>86557</td><td>银砂黑</td><td>37</td><td>19.78</td></tr>
<tr><td rowspan="3">2 月</td><td rowspan="2">福克斯</td><td>两厢 1.8L 自动时尚型</td><td>86L73</td><td>激情红</td><td>211</td><td>19.58</td></tr>
<tr><td>三厢 1.8L 自动时尚型</td><td>86L68</td><td>星光银</td><td>156</td><td>18.78</td></tr>
<tr><td>嘉年华</td><td>两厢 1.5L 自动时尚型</td><td>85778</td><td>烈焰红</td><td>140</td><td>19.38</td></tr>
<tr><td rowspan="14">3 月</td><td rowspan="8">蒙迪欧致胜</td><td>三厢 2.3L 自动时尚型</td><td>86526</td><td>银砂黑</td><td>234</td><td>17.98</td></tr>
<tr><td>GTDI 200 豪华型浅内饰</td><td>86532</td><td>银砂黑</td><td>237</td><td>17.70</td></tr>
<tr><td>GTDI 200 豪华型浅内饰</td><td>86560</td><td>银砂黑</td><td>94</td><td>15.68</td></tr>
<tr><td>GTDI 240 至尊型浅内饰</td><td>86538</td><td>钛晶灰</td><td>82</td><td>15.60</td></tr>
<tr><td>豪华型深内饰 2.3L</td><td>86555</td><td>银砂黑</td><td>0</td><td>14.60</td></tr>
<tr><td>三厢 2.3L 自动时尚型</td><td>86526</td><td>银砂黑</td><td>373</td><td>14.60</td></tr>
<tr><td>三厢 2.3L 自动时尚型</td><td>86526</td><td>钛晶灰</td><td>363</td><td>14.38</td></tr>
<tr><td>GTDI 200 时尚型</td><td>86557</td><td>银砂黑</td><td>4</td><td>14.09</td></tr>
<tr><td rowspan="3">福克斯</td><td>三厢 1.8L 自动豪华型</td><td>86L69</td><td>钛晶灰</td><td>358</td><td>16.19</td></tr>
<tr><td>两厢 1.5L 手动舒适型</td><td>86L71</td><td>旋舞橙</td><td>83</td><td>15.00</td></tr>
<tr><td>三厢 1.8L 手动舒适型</td><td>86L66</td><td>银砂黑</td><td>187</td><td>14.80</td></tr>
<tr><td rowspan="3">嘉年华</td><td>两厢 1.5L 自动运动型</td><td>85781</td><td>耀目黑</td><td>1</td><td>15.99</td></tr>
<tr><td>两厢 1.5L 自动运动型</td><td>85783</td><td>闪耀黄</td><td>10</td><td>15.19</td></tr>
<tr><td>三厢 1.5L 手动时尚型</td><td>85756</td><td>炫耀银</td><td>16</td><td>13.64</td></tr>
</table>

2-C2　请总结自动筛选与高级筛选的优缺点及应用范围。

学习任务页-3　数据的分组或分类

学生信息__________　　任务开始日期__________　　任务分值____15 分____

3-A1-1　何谓组？分组与分类的含义是什么？

3-A1-2　数据的分组需要遵循哪些原则？

3-A2-1　数值型数据的分组方法有哪些？

3-A2-2　单变量分组要点有哪些？

3-A2-3　组距式分组有哪两种类型？组距式分组要点有哪些？

3-A2-4　何谓组数、全距和组距？确定组数与组距的经验公式。

3-A2-5　何谓组中值？如何确定组中值？

3-A2-6　何为开口组？如何确定开口组的组距和组中值？

3-B1-1　某生产车间 50 名工人日加工零件数如下所示(单位：个)。

117	122	124	129	139	115	117	130	122	125
108	131	125	117	122	133	126	122	118	108
110	118	123	126	133	134	127	123	118	112
112	134	127	123	125	113	120	123	127	130
137	114	120	128	124	115	139	128	124	121

若采用单变量分组的方法，可以分成哪些组？

3-B1-2　如图 2-1 所示的是某电脑公司 2009 年前四个月各天的销售量数据(单位：台)。若采用组距式分组的方法，可以分成哪些组？

	A	B	C	D	E	F	G	H	I	J
1	某电脑公司2009年前四个月各天的销售量数据（单位：台）									
2	234	143	187	161	150	228	153	166	154	174
3	156	203	159	198	160	152	161	162	163	196
4	164	226	165	165	187	141	214	149	178	223
5	218	179	215	180	175	196	155	167	168	211
6	168	170	180	171	233	172	210	172	172	194
7	173	196	174	165	175	233	175	190	207	176
8	183	225	178	234	153	179	144	179	188	172
9	181	182	182	177	184	185	186	186	178	187
10	237	187	205	188	177	189	209	189	190	175
11	191	173	194	189	195	195	163	196	176	196
12	160	197	197	174	198	200	201	202	158	203
13	188	206	171	208	192	210	168	211	172	213

图 2-1　某电脑公司 2009 年前四个月各天的销售量数据

3-B2-1　请在“任务 3-B1-1”的基础上，把数据归类到对应的组。

3-B2-2　请在“任务 3-B1-2” 的基础上，把数据归类到对应的组。



学习任务页-4　数据展示：制表

学生信息__________　　任务开始日期__________　　任务分值　20 分　

4-A1-1　数据展示的意义有哪些？

4-A1-2　频数、频率、累计频数、累计频率的定义是什么？

4-A2-1　什么是频数分布表？制作频数分布表的意义有哪些？

4-A2-2　利用 Excel 制作频数分布表的方法有哪些？

4-A2-3　何谓数据透视表？透视表有哪些功能？

4-B1-1　请利用 Excel 制作“任务 3-B1-1”所列数据的频数分布表。

4-B1-2　请利用 Excel 制作“任务 3-B1-2”所列数据的频数分布表。

4-B1-3　现有 20 名学生的某门课成绩如下所示。

74	93	62	88	86	51	97	73	77	81
85	67	92	60	84	80	78	90	85	81

试为这些成绩数据编制频数分布表，并生成相应的直方图。

4-B1-4　在某高校在校男大学生中随机抽取 40 人，测得其身高如下所示(单位：cm)。

176	168	176	180	184	167	168	164	167	172
174	173	177	170	168	177	170	172	173	160
171	176	163	175	158	161	172	172	172	179
163	169	178	181	166	178	176	171	172	157

取组距为 5。试用 Excel 生成频数分布表和直方图。

4-B1-5　表 2-7 所示的是某学校数学建模学员的基本信息表。请根据所列数据，分别按系别、班级、生源地分类，并制作频数分布表。

表 2-7　某学校数学建模学员的基本信息表

序号	姓　名	性别	系　别	班　级	生源地
1	周若芳	女	工商系	工商 3	安徽
2	马新宇	男	工商系	连锁 1	安徽
3	耿冬冬	女	工商系	连锁 2	江苏
4	王华	女	工商系	连锁 3	山西
5	闫妮	女	工商系	连锁 3	辽宁
6	郭琳	女	工商系	市场 2	山东
7	李文	男	工商系	市场 2	山东
8	焦菲	女	工商系	市场 2	安徽
9	姗姗	女	工商系	市场 2	辽宁
10	青青	女	工商系	市场 2	安徽
11	香香	女	工商系	市场 2	安徽
12	晶晶	女	工商系	市场 2	安徽
13	冯雷	男	工商系	市场 2	吉林
14	宋康霖	男	工商系	市场 2	北京
15	赵大力	男	工商系	市场 2	北京
16	魏薇	女	工商系	市场 1	山西
17	徐胜楠	女	工商系	市场 1	安徽
18	刘翠青	女	工商系	市场 1	山西
19	金冰	男	工商系	市场 1	安徽
20	林爽	女	工商系	市场 1	辽宁
21	于敏	女	金融系	银行 1	北京
22	刘雪	男	金融系	银行 3	北京
23	陈宇宙	男	金融系	银行 3	内蒙古
24	李亚男	女	金融系	证券 1	辽宁
25	郭明阳	女	金融系	证券 1	吉林
26	黄启金	男	金融系	证券 1	江苏
27	高瑞霞	女	金融系	银行	河南
28	张杨	男	金融系	银行 1	北京
29	林伟	男	立信会计学院	会计 1	山东
30	孙婷	女	立信会计学院	会计 2	北京
31	常树林	男	信息物流	物管 1	内蒙古
32	徐一彤	男	信息物流	物管 2	北京
33	马永嘉	男	金融系	银行 2	北京

4-B2-1　请对“任务 4-B1-1”和“任务 4-B1-2”的结果进行合理的文字解释。

4-B2-2　请对“任务 4-B1-3”和“任务 4-B1-4”的结果进行合理的文字解释。

4-B2-3　请对“任务 4-B1-4”的结果进行合理的文字解释。

4-B2-4　请对“任务 4-B1-5”的结果进行合理的文字解释。

学习任务页-5　数据展示：绘图

学生信息__________　　任务开始日期__________　　任务分值____20 分____

5-A1　常用的数据展示图有哪些？绘制数据展示图的意义是什么？

5-A2-1　柱形图、条形图、折线图、饼图分别适合展示数据的哪些分布特点？

5-A2-2　何为散点图？用其来展示数据的什么特点？

5-A2-3　何为直方图？用其来展示数据的什么特点？

5-B1-1　请利用 Excel 绘制“任务 4-B1-1”的柱形图、条形图、饼图。

5-B1-2　请利用 Excel 绘制“任务 4-B1-2”的柱形图、条形图、饼图。

5-B1-3　请根据“任务 2-C1”所列数据，绘制汽车销售量随时间变化的折线图。

5-B1-4　表 2-8 所示的是 2000—2014 年北京市年用电量指标统计数据。请绘制“GDP与年用电量”“总人口与 GDP”的散点图。

表 2-8　2000—2014 年北京市年用电量指标统计数据

年份	总人口/万人	GDP/亿元	年用电量/万千瓦小时
2000	1 363.6	3 161.7	3 844 266
2001	1 385.1	3 708.0	3 999 415
2002	1 423.2	4 315.0	4 399 637
2003	1 456.4	5 007.2	4 676 056
2004	1 492.7	6 033.2	5 131 804
2005	1 538.0	6 969.5	5 705 364
2006	1 601.0	8 117.8	6 115 719

续表

年份	总人口/万人	GDP/亿元	年用电量/万千瓦小时
2007	1 676.0	9 846.8	6 670 089
2008	1 771.0	11 115.0	6 897 189
2009	1 860.0	12 153.0	7 391 465
2010	1 961.9	14 113.6	8 099 029
2011	2 018.6	16 251.9	8 217 055
2012	2 069.3	17 879.4	8 742 835
2013	2 114.8	19 800.8	9 131 113
2014	2 151.6	21 330.8	9 370 485

资料来源：北京市统计局 2014 统计年鉴，http://www.bjstats.gov.cn/.

5-B2　表 2-9 所示的是某学院 2010—2014 年在校生人数。请利用合适的数据展示图对该学院在校生的情况进行全面、详细地展示及分析。

表 2-9　某学院 2010—2014 年在校生人数　　单位：人

年　份	2010	2011	2012	2013	2014
男生	1 561	2 156	2 526	2 703	3 045
女生	1 071	1 407	2 226	2 883	3 305
合计	2 632	3 563	4 752	5 586	6 350

5-C1-1　请对“任务 5-B1-1”的结果进行合理的文字解释。

5-C1-2　请对“任务 5-B1-2”的结果进行合理的文字解释。

5-C1-3　请对“任务 5-B1-3”的结果进行合理的文字解释。

5-C1-4　请对“任务 5-B1-4”的结果进行合理的文字解释。

5-C1-5　请对“任务 5-B2”的结果进行合理的文字解释。

学习任务页-6　数据的统计描述

学生信息__________　　任务开始日期__________　　任务分值＿＿20 分＿＿

6-A1-1　可以从哪些方面对数据进行测定与描述?

6-A1-2　描述数据集中趋势，常用的描述统计量有哪些? 它们的定义是什么?

6-A1-3　描述数据离散程度，常用的描述统计量有哪些? 它们的定义是什么?

6-A2-1　平均数、众数、中位数分别表达什么含义?

6-A2-2　极差、方差、标准差、标准差系数分别表达什么含义？

6-B1-1　请在计算机上安装 Excel 数据分析程序。在 Excel 的“工具”菜单将出现“数据分析”命令选项。

6-B1-2　请利用 Excel 计算如表 2-1 所示数据中，语、数、外三门课程的描述统计量。

6-B1-3　随机抽取 10 人，测得其血清中的锌含量如下所示(单位：mol/L)。

16.8　24.2　17.4　10.4　15.8　12.1　16.6　20.4　17.1　19.6

试用 Excel 计算锌含量的中位数、样本均值、样本方差、标准差、标准差系数和极差等。

6-B1-4　请计算“任务 4-B1-4”所列数据的描述统计量。

6-B2-1　请对“任务 6-B1-2”的结果进行合理的文字解释。

6-B2-2　请对“任务 6-B1-3”的结果进行合理的文字解释。

6-B2-3　请对“任务 6-B1-4”的结果进行合理的文字解释。

6-C1　请按照上述数据处理方法与步骤，对表 2-1 所列数据进行完整分析，并写成一篇小论文。

学习任务页-zh1　公共自行车服务系统

（2013 年全国大学生数学建模竞赛 D 题节选）

学生信息__________　　任务开始日期__________　　任务分值　25 分

公共自行车服务作为一种低碳、环保、节能、健康的出行方式，正在全国许多城市迅速推广与普及。在公共自行车服务系统中，自行车租赁的站点位置及各站点自行车锁桩和自行车数量的配置，对系统的运行效率与用户的满意度有重要的影响。

浙江省温州市鹿城区公共自行车管理中心提供了其 20 天借车和还车的原始数据。讨论以下问题。

（1）分别统计各站点 20 天中每天及累计的借车频次和还车频次，并对所有站点按累计的借车频次和还车频次分别排序。另外，试统计分析每次用车时长的分布情况。

（2）试统计 20 天中各天使用公共自行车的不同借车卡（即借车人）数量，并统计数据中出现过的每张借车卡累计借车次数的分布情况。


学习任务页-zh2　储药柜的设计

（2014 年全国大学生数学建模竞赛 D 题节选）

学生信息__________　　任务开始日期__________　　任务分值　25 分　

储药柜的结构类似于书橱，通常由若干个横向隔板和竖向隔板将储药柜分割成若干个储药槽（见图 2-2）。为保证药品分拣的准确率，防止发药错误，一个储药槽内只能摆放同一种的药品。药品在储药槽中的排列方式如图 2-3 所示，药品从后端放入，从前端取出。实际储药柜中药品的摆放情况如图 2-4 所示。

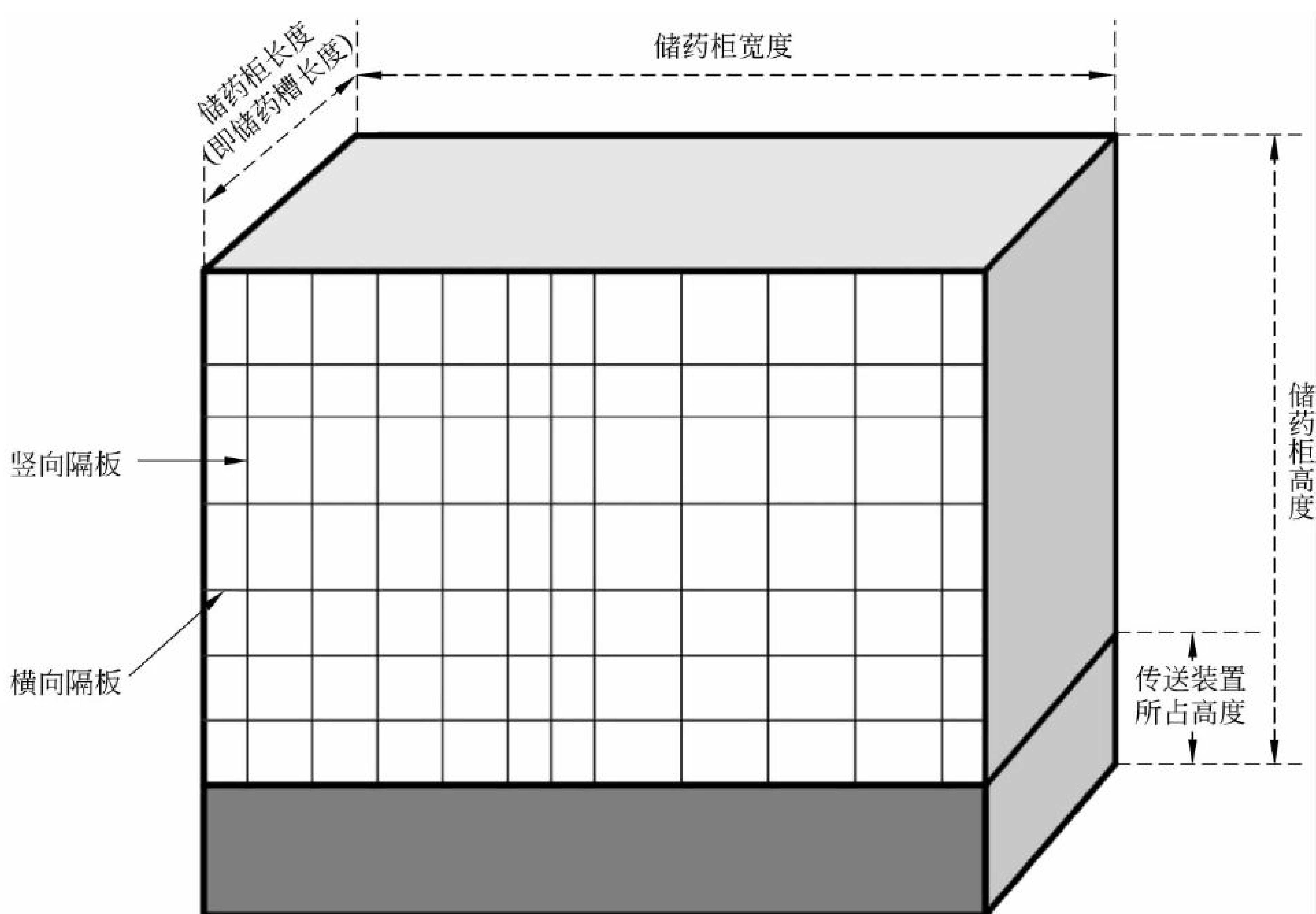


图 2-2　储药柜立体示意图

为保证药品在储药槽内顺利出入，要求药盒与两侧竖向隔板之间、与上下两层横向隔板之间应留 2mm 的间隙，同时要求药盒在储药槽内的推送过程中不会出现并排重叠、侧翻或水平旋转。在忽略横向和竖向隔板厚度的情况下，建立数学模型，给出下面两个问题的解决方案。

（1）药房内的盒装药品种类繁多，药盒尺寸规格差异较大，官网上给出了一些药盒的规格。请利用相关数据，给出竖向隔板间距类型最少的储药柜设计方案，包括类型的数量和每种类型对应的药盒规格。

（2）药盒与两侧竖向隔板之间的间隙超出 2mm 的部分可视为宽度冗余。增加竖向隔板的间距类型数量，可以有效地减少宽度冗余，但会增加储药柜的加工成本，同时降低储药槽的适应能力。设计时，希望总宽度冗余尽可能小，同时希望间距的类型数量尽可能少。仍利用相关数据，给出合理的竖向隔板间距类型的数量，以及每种类型对应的药品编号。

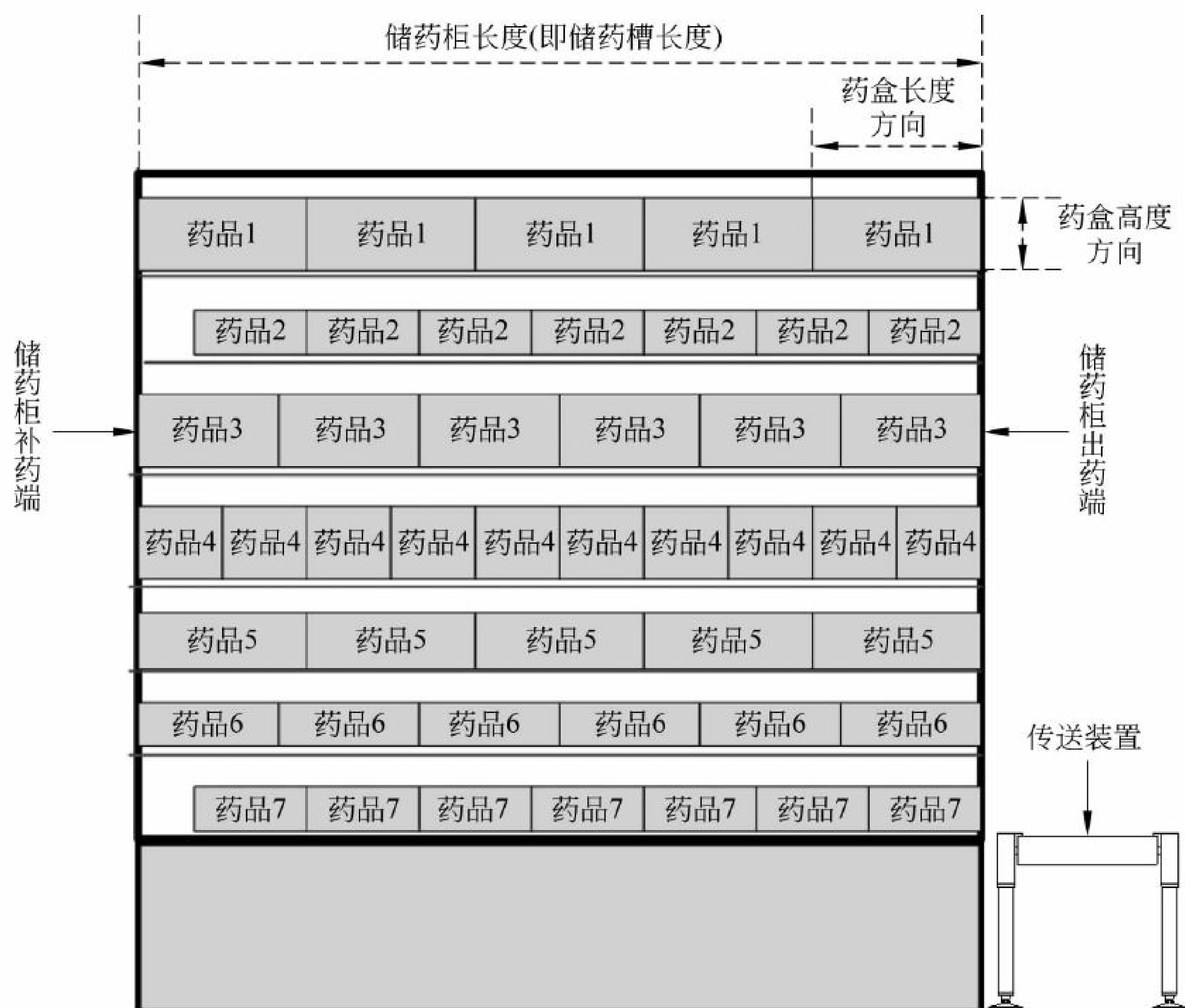


图 2-3　储药柜的侧剖面及药品摆放示意图



图 2-4　储药槽药品摆放情况

学习信息库1——数据预处理概述

数据无处不在，也是各行各业各部门的重要资料，但如果不对数据进行预处理，合理地显示与描述的话，一大堆数据将没有任何价值。本章主要介绍整理、显示与描述数据的方法，使错综复杂的数据变成有价值的信息，为各种客观现象或事物的决策或判断提供保障。这也是数据处理的目的。

1. 数据的概念

数据就是人们通过观察统计、调查统计、实验统计等方式得到的各种现象的信息。这个信息可以是数字、文字、图表、音频、视频，等等。

正确、完整的数据能够反映一种现象或事物的本质，这就是人们热衷于收集数据、分析数据的原因，以便达到“透过现象看本质”的目的。

2. 数据的分类

数据根据标志不同，分成很多类。比如，按样式，分为数值型数据和属性型数据；按变化，分为静态数据和动态数据；按时间，分为时间序列数据、截面数据和面板数据；按来源，分为原始数据和二手数据。这里主要介绍数值型数据和属性型数据。

（1）数值型数据

数值型数据又分为连续型数据和离散型数据。数值型数据说明的是现象的数量特征，通常用数值来表示。比如，人的正常体温为36.5℃，某人的身高为170cm，上证指数为3 600点，某运动员获得10个篮板、25分等。

连续型数据是能够连续取值的数据，而离散性数据是不能够连续取值的数据，只能取到若干个分散的值。一般情况下，取值都是整数。比如，36.5℃、170cm、3 600点等都是连续型数据，而10个篮板、25分等是离散型数据。

（2）属性型数据

属性型数据又称品质数据或分类数据，它说明的是事物的品质特征，通常用文字来表示。比如，人分为男、女两个性别，文化程度分为小学、初中、高中、大学、研究生、博士六个程度，年度考核成绩分为不及格、及格、优秀三个等级。

3. 数据的收集

数据的收集是数据处理与分析的基础阶段，需要在一般情况下，由专业的人员或机构运用科学的方法和手段收集客观数据。数据的客观性和可重复性是数据收集的基本原则。原始数据不能随意更改。若违背这些原则，可能造成对实际情况的错误判断或错误决策。关于数据收集更多的知识，这里不再详细介绍，本书主要介绍如何处理已有数据。

4. 数据预处理的基本步骤

对已有数据的预处理一般遵循下述基本步骤，也可按照需求灵活取舍。

1）数据的审核、筛选、排序

数据的审核、筛选和排序是数据预处理的第一步，是对数据分类或分组的前提和必不可少的步骤。

(1) 数据的审核。审核数据的目的是保证数据的质量。审核工作贯穿于数据处理的全过程。

对于原始数据,应从完整性、准确性两方面进行审核。完整性审核主要是检查数据是否有遗漏,各项数据是否齐全等;准确性审核包括计算检查和逻辑检查。

计算检查是从定量角度对数据进行审核,检查计算结果和计算方法是否有错误。例如,各项数字之和是否等于相应的合计数;各组结构比例之和是否等于 1 或 100%;出现在不同表格上的同一指标值是否一致等。该项检查主要用于数值型数据的审核。

逻辑检查是从理论的角度审核数据是否符合逻辑,是否说得通。比如,在家庭住户调查登记中,填写内容为:某人 9 岁,文化程度为研究生,职业为经理。这显然不符合逻辑,在登记中有误,应更正。

数据在经过审核后,可进一步加工处理。对于审核过程中发现的错误,应尽可能纠正;如果无法弥补,要对数据进行筛选。

(2) 数据的筛选。数据筛选包括以下两方面内容:一是剔除某些不符合要求的数据或明显错误的数据;二是筛选出符合某种特定条件的数据,剔除不符合特定条件的数据。

大量数据的筛选很难手动完成,应借助计算机软件。常用的软件有 Excel,具体的操作方法在后面详细介绍。

(3) 数据的排序。数据排序是指按一定顺序将数据排列,以便通过浏览数据发现一些明显的特征或趋势,找到解决问题的线索。此外,排序有助于对数据进行检查审核、重新归类或分组等。

对于不同类型的数据,其排序的依据和标志不同。数值型数据可根据数值的大小,升序或降序排列;字母型数据可按字母顺序排列;汉字型数据可按首字母或笔画排序。数据排序可借助 Excel 轻松完成。

2) 分组或分类

分组或分类是在数据预处理的基础上,根据研究的目的,按照数据资料的性质与特点,对数据进行分组归类;或利用某种标志,将数据资料划分为若干组成部分,这些组成部分称为“组”。

针对不同类型的数据,采用的处理方式和方法不同。

分类数据本身就是对数据的一种分类。因此,若没有别的需要,无须进一步分类。下面主要介绍数值型数据分组汇总的基础知识与方法。

(1) 分组的原则。数据分组必须遵循两个原则:完整性原则和互斥性原则。

完整性原则:数据资料的全部数据都有所归属,即原始数据的每一条数据或每一个数据都归属于某一个组。

互斥性原则:数据资料的每一条数据或每一个数据只能归属于某一组,而不是既可以归这个组又可以归那个组。

(2) 分组的方法。数值型数据分组的方法有单变量分组和组距式分组两种。

① 单变量分组是指将一个变量值作为一组,一般适合于离散型变量,且变量取值范围不大的数据。

单变量分组要点:将一个变量值作为一组;适合于离散型数据;适合于变量值较少的数据。

单变量分组示例如表 2-10 所示。

表 2-10　定点投篮成绩分组表

投中数量分组/个	人数/人	投中数量分组/个	人数/人
20	2	16	9
19	3	15	4
18	3	14	1
17	8	合计	30

② 组距式分组是指将变量按取值顺序划分为几个区间，把属于一个区间的变量值归为一组，即有几个区间就有几个组。区间的长度称为组距。

组距式分组又分为等距分组和不等距分组（异距分组）。

等距分组是指每个组的组距相等。一般情况下采用等距分组方式，尤其是在变量值变动比较均匀的情况下，适合采用等距分组方式。等距分组便于执行各种操作，如计算、制表、绘图等。

不等距分组是指每个组的组距不相等，一般适用于变量变动很不均匀的数据。

组距式分组的要点：将变量值的一个区间作为一组；适合于连续变量；适合于变量值较多的情况；遵循“不重不漏”的原则。

组距分组示例如表 2-11 所示。

表 2-11　学生数学成绩分组表

成绩分组（满分：100 分）/分	学生人数/人	成绩分组（满分：100 分）/分	学生人数/人
60 以下	5	81～90	10
61～70	13	91～100	7
71～80	15	合　计	50

组距式分组的几个基本概念。

a. 组数：分组的个数。确定组数可采用斯特杰斯经验公式，这是美国学者斯特杰斯(Sturges)创建的。其公式为

$$k = 1 + \frac{\lg n}{\lg 2} = 1 + 3.322\lg n \tag{2-1}$$

其中，k 为组数；n 为数据个数。

b. 全距：全局最大值与最小值之差，即

$$R = x_{\max} - x_{\min} \tag{2-2}$$

其中，R 为全距；$x_{\max}$为全局最大值；$x_{\min}$为全局最小值。

c. 组距：区间的长度，即区间右端点减去区间左端点所得的值。组数确定以后，组距自然被确定，其公式为

$$d = \frac{R}{k} \tag{2-3}$$

其中，d 为组距；R 为全距；k 为组数。

注意：上述组距公式和组数公式仅供参考，在实际问题中不能完全按照公式确定，应灵活掌握。

d. 组中值：每组上限（区间右端点）与下限（区间左端点）之间的重点数值，其计算公式为

$$x_i = \frac{a_i + b_i}{2} \tag{2-4}$$

其中，x_i 为每组的组中值；a_i 为每组的上限；b_i 为每组的下限。

e. 开口组的组距和组中值：在组距分组中，如果全局最大值和最小值与其他数据相差悬殊，为避免出现空白组（即没有变量值的组）或个别极端值被漏掉，第一组和最后一组可以采取“* * 以下”或“* * 以上”形式的分组，称为开口组。开口组的组距可以其相邻组的组距确定。比如，表 2-7 中，学生数学成绩分为 60 分以下、61～70 分、71～80 分、81～90 分、90～100 分，则 60 分以下组的组距可以等于其相邻组 61～70 分的组距 9，即把这一组看成 51～60 分，其组中值是(60+51)/2=55.5(分)。

组距式分组的基本步骤：确定组数；确定组距；统计出各组的频数，并整理成频数分布表。

注意：频数及频数分布表的概念在 2.2.3 小节介绍。

例 2-1 某生产车间的 50 名工人日加工零件数（单位：个）如下所示。

117	122	124	129	139	115	117	130	122	125
108	131	125	117	122	133	126	122	118	108
110	118	123	126	133	134	127	123	118	112
112	134	127	123	125	113	120	123	127	130
137	114	120	128	124	115	139	128	124	121

试采用组距式分组方式对数据分组。

分析：一组数据分多少组合适呢？一般与数据本身的特点以及数据的多少有关。由于分组的目的之一是为了观察数据分布的特征，因此组数应适中。组数太少，数据分布过于集中；组数太多，数据分布过于分散，都不便于观察数据分布的特征和规律。组数确定后，组距即确定。

第一步：确定组数。

可根据式(2-1)确定。本题中，数据个数 $n=50$，则

$$k = 1 + 3.322 \times \lg 50 \approx 7$$

对结果四舍五入取整，即数据分成 7 组。

第二步：确定组距。

根据式(2-2)和式(2-3)确定。本题中，

$$R = x_{\max} - x_{\min} = 139 - 107 = 32, \quad d = \frac{R}{k} = \frac{32}{7} \approx 5$$

为便于计算，组距取 5 或 10 的倍数，而且第一组的下限应低于最小变量值，最后一组的上限应高于最大变量值，因此组距取 5。

第三步：统计出各组的频数并整理成频数分布表。比如，对上述数据分组，得到如

表 2-12 所示的频数分布表。

表 2-12　某车间 50 名工人日加工零件数分组表

按零件数分组/个	频数/人	频率/%	按零件数分组/个	频数/人	频率/%
105～110	3	6	125～130	10	20
110～115	5	10	130～135	6	12
115～120	8	16	135～140	4	8
120～125	14	28	合　计	50	100

采用组距分组时,需要遵循“不重不漏”的原则。“不重”是指一项数据不能同属两个不同的组;“不漏”是指组别能够穷尽,即所有数据项都有组可归,不能遗漏。

对于连续型变量,为解决“不重”的问题,统计分组时习惯上规定“上组限不在内”,即当相邻两组的上、下限重叠时,恰好等于某一组上限的变量值不算在本组内,而计算在下一组内。例如,在表的分组中,“125”这一数值不计算在“120～125”这一组内,而计算在“125～130”这一组,其余类推。

对于离散型变量,可以采用相邻两组组限间断的办法解决“不重”的问题。例如,对上述数据做如表 2-13 所示的分组。

表 2-13　某车间 50 名工人日加工零件数分组表

按零件数分组/个	频数/人	频率/%	按零件数分组/个	频数/人	频率/%
105～109	3	6	125～129	10	20
110～114	5	10	130～134	6	12
115～119	8	16	135～140	4	8
120～124	14	28	合　计	50	100

3）数据展示：制表

数据本身往往会信息多,数量大,需要用精简的方式展示出来,以便找到某种规律、特征、趋势等。数据可以采用表格或图形来展示,但各有利弊,图、表结合,效果更佳。

表格适宜展示数据的细节,对于数据的查询、处理和引用较方便,但不能形象、直观地展示数据的总体特征,而图形展示能弥补这一缺点。用图形展示数据,直观、易懂、形象,能够较好地反映数据的总体特征、各部分的比例关系或差别,以及数据的发展趋势。

(1) 基本概念

① 频数：也称次数,用 f_i 表示,是指原始数据被分配到各组中的个数。

② 频率：也称比率或比重,是指各组频数与总数之比,即

$$F_i = \frac{f_i}{\sum f_i} \tag{2-5}$$

其中,F_i 为各组的频率;f_i 为各组的频数; $\sum f_i$ 为各组频数总和。

③ 累计频数(频率)：将各组的频数(频率)累加起来的总和。

为了数据分析的需要,有时需要观察某一数值以下或某一数值以上的频数或频率之和,

这就是计算累计频数或频率的意义所在。

④ 频数分布表：在分组或分类的基础上，将原始数据逐个分配到不同的组内，计算各组的频数与频率，进而研究数据的分布特征。于是构造出频数分布表，见表 2-7～表 2-9。

(2) 制作频数分布表的意义

- 利用频数分布表，可以划分现象的类别，从而对数据及其特征有初步的了解。
- 利用频数分布表，可以计算出各类别的频率，从而反映现象的内部结构特征。
- 频数分布表也是制作各种图的基础，一般先有表后有图。

(3) 频数分布表的制作方法

频数分布表的制作借助 Excel 可轻松完成。下面介绍两种制作频数分布表的方法。

① 利用 Excel 数据透视表制作频数分布表。

数据透视表是一种对大量数据进行快速汇总和交叉显示数据信息的交互式工具。它可以动态地改变报表的版面布置，以便按照不同的方式分析数据，发现数据的本质特点。之所以称为透视表，因为通过它，不仅能看到表面的数据，更能透析隐藏的数据信息。当数据源中的数据成千上万时，使用数据透视表处理数据应是首选方案。

② 利用 Excel 的"数据分析"工具制作频数分布表。

Excel 有一组功能强大的数据分析工具库，用来完成频数分布表、直方图、相关分析、回归分析、描述统计等统计分析。默认状态下，一般不显示分析工具库，需要手动加载，加载方法如下所述。

在 Excel 页面中依次单击"文件"→"选项"→"加载项"→"分析工具库"→"转到…"→"分析工具库"→"确定"，完成 Excel 数据分析程序的安装，在 Excel 的"数据"选项卡中出现"数据分析"选项卡。该选项卡中包括可实现上述统计分析功能的选项。具体操作如图 2-5～图 2-9 所示。



图 2-5 依次单击"文件"→"选项"



图 2-6　单击“加载项”



图 2-7　依次单击“分析工具库”→“转到”

4）数据展示：绘图

在表格的基础上，可以用各种图形把表格中的数据形象地展示出来。常用的数据展示图有柱形图、条形图、饼图、散点图、折线图、直方图等。绘制图形可以在频数分布表的基础上，利用 Excel 图表向导完成。

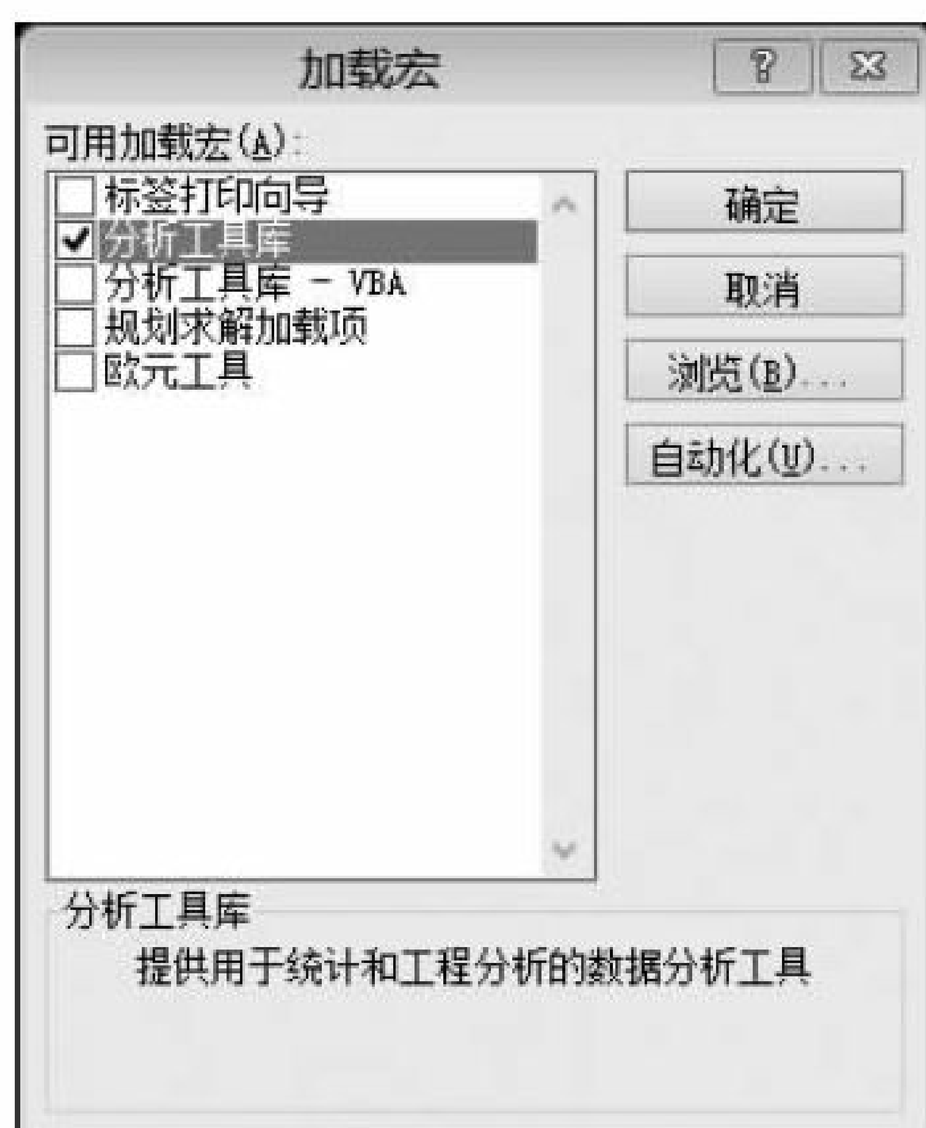


图 2-8　依次单击“分析工具库”→“确定”

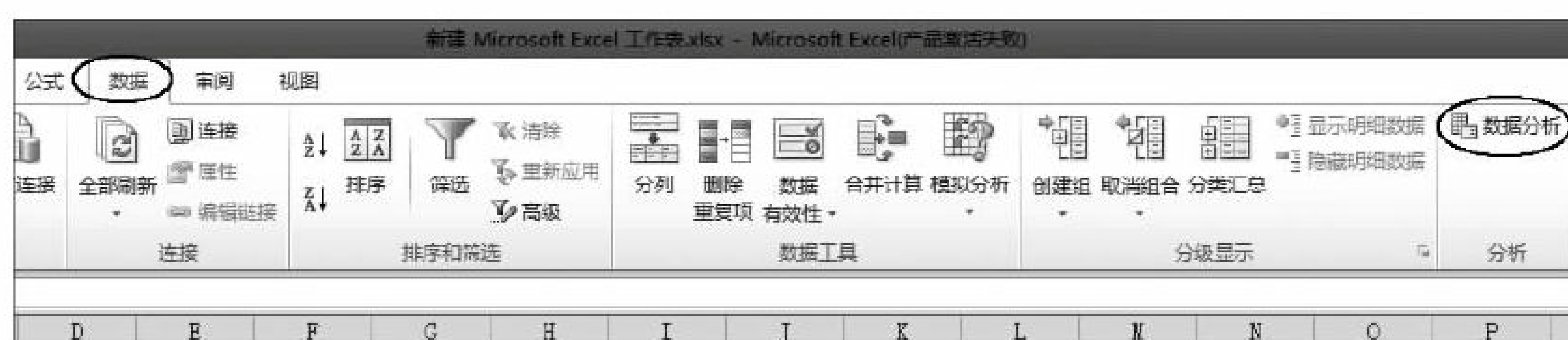


图 2-9　“数据”选项卡下的“数据分析”选项

(1) 柱形图：是最常见的数据图形之一，用来展示不同项目之间的数量关系或差别，可以直观、形象地显示各组的频数情况。

(2) 条形图：也是常见的数据图形之一，它的作用类似于柱形图，区别是：柱形图的柱是“竖”的，条形图的柱是“横”的。可根据不同的数据灵活选用柱形图或条形图。

(3) 饼图：用于展示数据中不同类别在总体中所占的比重，用于反映现象的结构及各类别之间的比例。

(4) 散点图：用于展示两个变量之间的相关关系。图形的横轴表示一个变量，纵轴表示另一个变量。相关分析与回归分析经常用到散点图。

(5) 折线图：主要用于展示数据变动的方向和趋势，表现数据在不同时期发展变化的趋势，适用于时间序列数据。

(6) 直方图：主要用于展示数据中各类别数据出现的频数和频率，显示现象中不同类别的数量分布形态，适用于数值型数据。绘制直方图需要用到 Excel 的“数据分析”功能。

5) 数据的统计描述

数据经过整理与展示后，人们大致了解了数据分布的类型和特点，但缺少代表性的数量特征，无法准确地描述数据分布的集中趋势和离散程度。

(1) 集中趋势。集中趋势反映数据向其中心值靠拢或聚集的趋势，其主要指标有平均数、众数和中位数，用于综合反映数据的平均水平或集中趋势。

① 平均数：总体数据的平均化，反映总体数据的平均水平。平均数分为算数平均数、调和平均数与几何平均数等。这里主要介绍算数平均数。

简单算数平均数：通常意义下的均值，即全体数值相加再除以数值个数，适用于未分组

的数据，其公式如下所示。

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{2-6}$$

其中，$\bar{x}$ 为算术平均数；x_i 为每一个数值；n 为数据个数。

加权算数平均数：适用于已经分组的数据，且已知频数分布的情况。它不是简单意义下的全体数值相加，而是每个数值乘以对应的频数或频率后做平均，其公式如下所示。

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{\sum_{i=1}^{n} f_i} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{f_1 + f_2 + \cdots + f_n} \tag{2-7}$$

或

$$\bar{x} = \sum_{i=1}^{n} x_i \cdot F_i = x_1 F_1 + x_2 F_2 + \cdots + x_n F_n \tag{2-8}$$

其中，$\bar{x}$ 为加权算术平均数；x_i 为每一个数值或组距式分组数据每组的组中值；f_i 为每组的频数；F_i 为每组的频率。这里，频数或频率就是权数。

简单算数平均数，意味着每个数值在总体数据中的作用都是一样的；而加权算数平均数，意味着每个数值在总体数据中的作用是不一样的。通俗地说，对总体数据影响大的数值，权重会大一些，反之小一些。

用平均数描述数据的优缺点如下所述。

优点：平均数会考虑到每个数值的变动，且一组数据只有一个均值，所以便于比较不同组数据之间的一般水平，也便于比较同一组数据在不同时期的平均水平。

缺点：它会受到不能真正代表一组数据的极端值的影响。

② 众数：是指一组数据中出现频数或频率最多的数量值。

用众数描述数据的优缺点如下所述。

优点：不受极端值的影响。若数据出现较多极端值，可用众数代表其一般水平。

缺点：当一组数据没有重复值出现，集中趋势不明显时，众数不存在；而当有些数据重复出现的次数相同时，会有多个众数。

③ 中位数：是指一组数据中按照从大到小或从小到大的顺序排列后，位于数列中点位置的数值。

用中位数描述数据的优缺点如下所述。

优点：受极端值的影响较小，适用于描述定序数据的一般水平。比如，对于用等级、名次等表示的数据，可采用中位数来代表一般水平。

缺点：没有考虑到所有数据的影响，仅是一种大致的集中趋势指标。

（2）离散程度。描述集中趋势的指标是用一个数值反映总体数据的一般水平或集中趋势，而离散程度反映数据远离其中心值的程度。描述离散程度的主要指标有极差、方差、标准差和标准差系数，用于综合反映数据的离散程度。

离散指标可用于衡量平均指标的代表性。平均指标能在多大程度上代表总体数据的整体水平，取决于总体数据的离散程度。离散程度越大，平均指标的代表性越小；离散程度越小，平均指标的代表性越大。

① 极差：也称全距，是指全局最大值与最小值之差。与全距的概念一致，如式(2-2)所示。

极差是最简单的离散指标，用于反映数据变动范围的大小很直观且易懂，但它只考虑到两个极端值的影响，没有考虑中间数据的影响。可初步认为，极差越大，数值变动范围越大，均值的代表性较差；反之，极差越小，数值变动范围越小，均值代表性较好。

② 方差：是指各数值与其平均数之差的平方的平均，常用 σ^2 表示，其公式如下所示。

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - x)^2}{n} \tag{2-9}$$

③ 标准差：是方差的平方根，常用 σ 表示，其公式如下所示。

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x)^2}{n}} \tag{2-10}$$

标准差是衡量数据离散程度最灵敏的指标。标准差越大，数值变动范围越大，均值的代表性较差；反之，标准差越小，数值变动范围越小，均值代表性较好。

④ 标准差系数：是一组数据的标准差与其均值的比值，其公式如下所示。

$$\bar{\sigma} = \frac{\sigma}{\bar{x}} \tag{2-11}$$

其中，$\bar{\sigma}$ 为标准差系数；σ 为标准差；$\bar{x}$ 为平均数。

极差、方差、标准差这三个指标都带着原始数据的计量单位。不同计量单位的数据之间无法直接比较，因此有必要进行无量纲化处理。标准差系数就是无量纲数，即没有计量单位，适用于对比分析均值水平不同或计量单位不同的两组数据的离散程度大小。标准差系数大，说明该组数据的离散程度大；标准差系数小，说明该组数据的离散程度小。

利用 Excel 的数据分析功能可一次性地算出以上各描述统计量，因此关于它们具体的计算方法，这里不详细介绍。

6）形成文字解释

数据经过上述几个步骤的预处理后，会得到数据展示表、数据展示图及数据的各种描述统计量。这些结果只是图表和数字，数据处理的真正目的是要把这些信息转化成人人能够看懂的文字解释。如何才能做到这一点呢？首先，要理解各种图表及描述统计量的具体含义；其次，在此基础上多做练习，学习信息库 2 例题后的“总结”部分就是对每个数据处理结果的简单的文字解释。

学习信息库 2——利用 Excel 做数据预处理

Excel 拥有简便、有效的数据预处理功能。下面详细介绍数据的筛选、排序、数据透视表、FREQUENCY 函数和数据展示图等功能。

1. 数据的筛选

Excel 中有自动筛选和高级筛选两个功能。自动筛选只能针对单一条件进行数据筛选，高级筛选可以一次性地针对多个条件进行筛选。

（1）自动筛选

自动筛选的步骤如下。

第一步：确定筛选区域，用鼠标拖曳选中即可。

第二步：依次单击"数据"→"筛选"，工作表每列上出现倒三角按钮。

第三步：设置筛选条件。单击倒三角按钮，然后在下拉列表中设置筛选条件。

第四步：单击"确定"按钮。

(2) 高级筛选

高级筛选的步骤如下。

第一步：在 Excel 表格空白处输入高级筛选条件。

第二步：依次单击"数据"→"高级筛选"按钮，打开"高级筛选"对话框。

第三步：设置"方式"，输入"列表区域"和"条件区域"，然后单击"确定"按钮，筛选出满足条件的数据。

例 2-2 图 2-10 所示的是一个班 20 名学生的高考成绩信息截图。请完成以下三个任务。

任务 1：筛选出数学成绩前 5 名的学生。

任务 2：筛选出总成绩大于 380 分的学生。

任务 3：筛选出三门课程成绩都大于 60 分的学生。

分析：任务 1 和任务 2 是单一条件筛选，可以用自动筛选功能完成；任务 3 是多项条件筛选，可以用高级筛选功能完成。具体步骤如下。

第一步：确定筛选区域 $A$1：$E$21，用鼠标拖曳选中。

第二步：依次单击"数据"→"筛选"，工作表每列上出现倒三角按钮，如图 2-11 所示。

	A	B	C	D	E
1	姓名	总成绩	语文	数学	英语
2	丁骧	341.00	100.00	73.00	51.00
3	郭楠	374.00	95.00	85.00	57.00
4	李璐	327.00	83.00	76.00	65.00
5	李铭	356.00	91.00	66.00	62.00
6	刘强	369.00	99.00	83.00	42.00
7	龙杨	276.00	85.00	51.00	61.00
8	思思	307.00	93.00	74.00	55.00
9	孙灵	324.00	97.00	73.00	56.00
10	孙淼	324.00	86.00	73.00	71.00
11	王芳	281.00	93.00	115.00	73.00
12	王菁	375.00	98.00	74.00	49.00
13	王瑞	347.00	105.00	30.00	76.00
14	王婷	415.00	107.00	69.00	48.00
15	王雪	384.00	91.00	102.00	48.00
16	吴源源	299.00	91.00	59.00	60.00
17	邢文	357.00	105.00	69.00	61.00
18	闫冬	404.00	105.00	75.00	45.00
19	杨新	337.00	86.00	94.00	42.00
20	张欧	383.00	97.00	59.00	82.00
21	赵静	353.00	91.00	78.00	48.00

图 2-10 20 名学生的高考成绩信息截图

	A	B	C	D	E
1	姓名	总成绩	语文	数学	英语
2	丁骧	341.00	100.00	73.00	51.00
3	郭楠	374.00	95.00	85.00	57.00
4	李璐	327.00	83.00	76.00	65.00
5	李铭	356.00	91.00	66.00	62.00
6	刘强	369.00	99.00	83.00	42.00
7	龙杨	276.00	85.00	51.00	61.00
8	思思	307.00	93.00	74.00	55.00
9	孙灵	324.00	97.00	73.00	56.00
10	孙淼	324.00	86.00	73.00	71.00
11	王芳	281.00	93.00	115.00	73.00
12	王菁	375.00	98.00	74.00	49.00
13	王瑞	347.00	105.00	30.00	76.00
14	王婷	415.00	107.00	69.00	48.00
15	王雪	384.00	91.00	102.00	48.00
16	吴源源	299.00	91.00	59.00	60.00
17	邢文	357.00	105.00	69.00	61.00
18	闫冬	404.00	105.00	75.00	45.00
19	杨新	337.00	86.00	94.00	42.00
20	张欧	383.00	97.00	59.00	82.00
21	赵静	353.00	91.00	78.00	48.00

图 2-11 出现倒三角按钮的数据截图

第三步：单击倒三角按钮，然后在下拉列表中设置筛选条件。

单击"数学"后的倒三角按钮，选择"降序"排列，取其前 5 名，即完成任务 1，筛选结果如图 2-12 所示。这时，"数学"后面的倒三角按钮后面出现一个向下的箭头，表明对这一列进行过降序筛选。

	A	B	C	D	E
1	姓名	总成绩	语文	数学	英语
2	王芳	281.00	93.00	115.00	73.00
3	王雪	384.00	91.00	102.00	48.00
4	杨新	337.00	86.00	94.00	42.00
5	郭楠	374.00	95.00	85.00	57.00
6	刘强	369.00	99.00	83.00	42.00

图 2-12 数学成绩前 5 名的筛选结果截图

总结：从筛选结果可知，数学成绩前 5 名

的学生分别是王芳、王雪、杨新、郭楠和刘强。其中，王芳的数学成绩最好。

单击“总成绩”后的倒三角按钮，选择“数字筛选”，然后单击“大于”下拉列表，并输入“380”，如图 2-13 所示，单击“确定”按钮，完成任务 2 的筛选，其结果如图 2-14 所示。

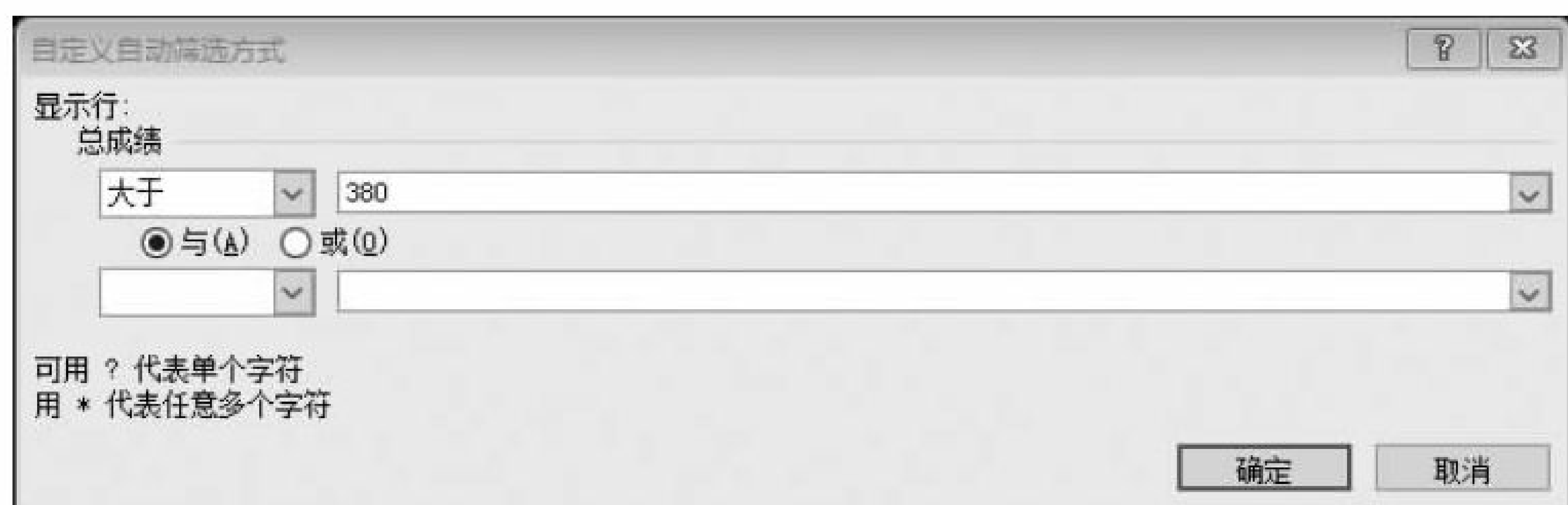


图 2-13 “大于”选项窗口

“数字筛选”选项里有各种筛选条件，比如“大于”“小于”“大于等于”“小于等于”等，可根据需要选用。

总结：从筛选结果可知，总成绩大于380 分的学生共有 4 名，分别为王雪、闫冬、王婷和张鸥。其中，王婷的总成绩最高，为 415 分。

任务 3 的操作步骤如下所述。

第一步：在 Excel 表空白处输入高级筛选条件。比如“语文＞60 分，数学＞60 分，英语＞60 分”，如图 2-15 所示。

第二步：依次单击“数据”→“高级筛选”按钮，打开“高级筛选”对话框，如图 2-16 所示。

	A	B	C	D	E
1	姓名	总成绩	语文	数学	英语
3	王雪	384.00	91.00	102.00	48.00
9	闫冬	404.00	105.00	75.00	45.00
15	王婷	415.00	107.00	69.00	48.00
19	张欧	383.00	97.00	59.00	82.00

图 2-14 总成绩大于 380 分的筛选结果截图

语文	数学	英语
>60分	>60分	>60分

图 2-15 高级筛选条件截图

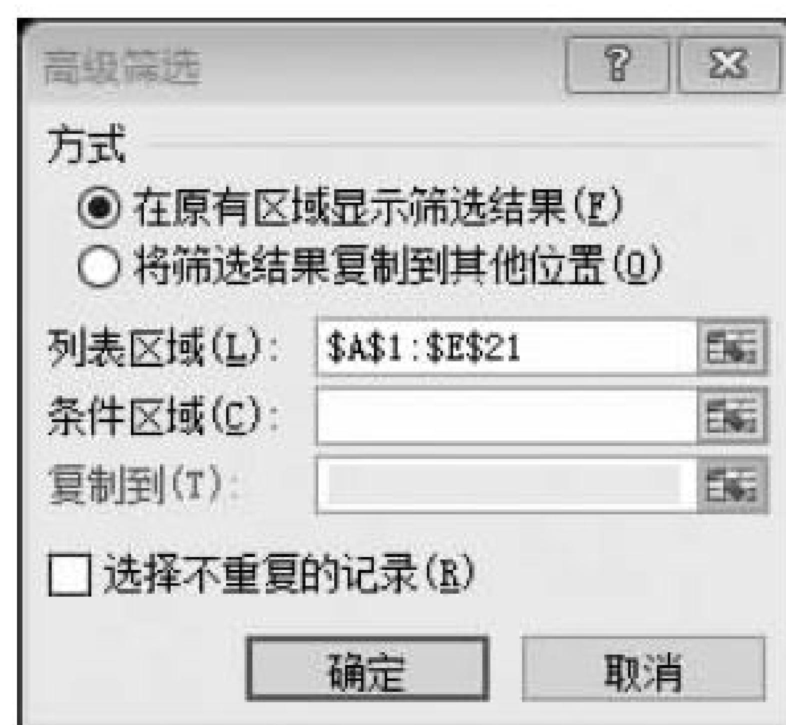


图 2-16 “高级筛选”对话框

第三步：设置“方式”，选择“列表区域”和“条件区域”，然后单击“确定”按钮，筛选出满足条件的数据，如图 2-17 所示。

姓名	总成绩	语文	数学	英语
王芳	281.00	93.00	115.00	73.00
李璐	327.00	83.00	76.00	65.00
孙淼	324.00	86.00	73.00	71.00
邢文	357.00	105.00	69.00	61.00
李铭	356.00	91.00	66.00	62.00

图 2-17 高级筛选结果截图

总结：从筛选结果可知，语文、数学、英语成绩都在 60 分以上的学生有 5 人，分别为王芳、李璐、孙森、邢文和李铭。

2. 数据的排序

利用 Excel 进行数据排序的步骤如下。

第一步：选中数据。

第二步：依次单击“数据”→“排序”按钮，打开“排序”对话框。

第三步：在“排序”对话框中进行相应的设置，并单击“确定”按钮。

例 2-3 按下列要求对例 2-2 中的数据进行排序。

(1) 分别对总成绩和数学成绩降序排列，对语文和英语成绩升序排列。

(2) 对姓名首字母升序排列。

下面以总成绩的降序排序为例，介绍 Excel 的排序方法，其他排序类似。其步骤如下。

第一步：确定筛选区域 $A$1:$E$21，用鼠标拖曳选中。

第二步：依次单击“数据”→“排序”按钮，打开“排序”对话框，如图 2-18 所示。



图 2-18 “排序”对话框

第三步：单击“主要关键字”下拉按钮，然后选择“总成绩”；在“排序依据”框中选择“数值”，在“次序”框中选择“降序”，然后单击“确定”按钮，得到排序结果，如图 2-19 所示。

	A	B	C	D	E
1	姓名	总成绩	语文	数学	英语
2	王芳	281.00	93.00	115.00	73.00
3	王雪	384.00	91.00	102.00	48.00
4	杨新	337.00	86.00	94.00	42.00
5	郭楠	374.00	95.00	85.00	57.00
6	刘强	369.00	99.00	83.00	42.00
7	赵静	353.00	91.00	78.00	48.00
8	李璐	327.00	83.00	76.00	65.00
9	闫冬	404.00	105.00	75.00	45.00
10	思思	307.00	93.00	74.00	55.00
11	王菁	375.00	98.00	74.00	49.00
12	丁骥	341.00	100.00	73.00	51.00
13	孙灵	324.00	97.00	73.00	56.00
14	孙淼	324.00	86.00	73.00	71.00
15	王婷	415.00	107.00	69.00	48.00
16	邢文	357.00	105.00	69.00	61.00
17	李铭	356.00	91.00	66.00	62.00
18	吴源源	299.00	91.00	59.00	60.00
19	张欧	383.00	97.00	59.00	82.00
20	龙杨	276.00	85.00	51.00	61.00
21	王瑞	347.00	105.00	30.00	76.00

图 2-19 “总成绩”降序排列结果

3. 利用数据透视表制作频数分布表

利用数据透视表制作频数分布表的步骤如下所述。

第一步：选中数据源(无须全选数据，把鼠标定位在其中一个单元格即可)。

第二步：依次单击“插入”→“数据透视表”，建立一个空白的数据透视表。

空白透视表由两部分组成，即空白数据透视表(见图 2-20)和数据透视表字段列表(见图 2-21)。

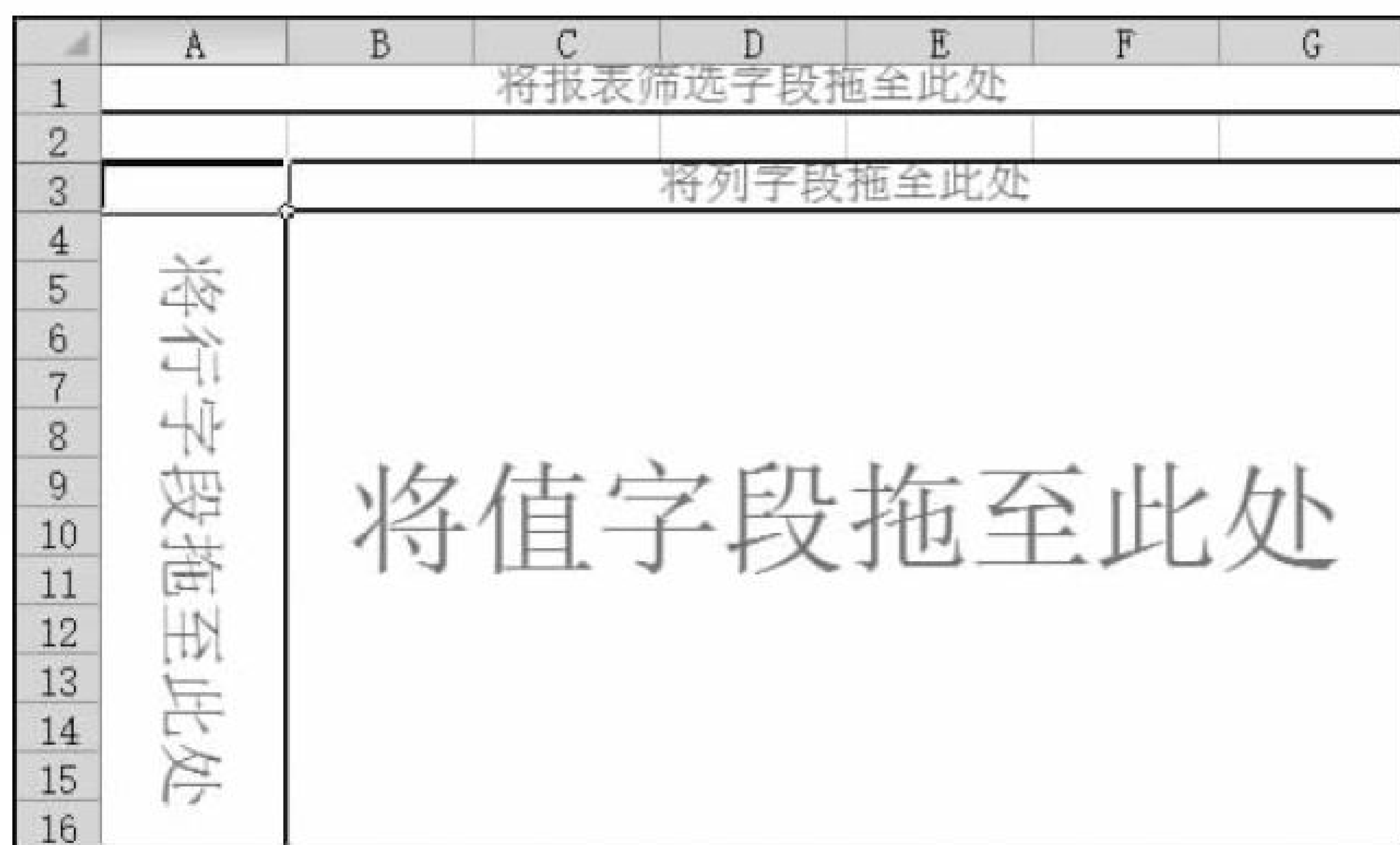


图 2-20　空白数据透视表



图 2-21　数据透视表字段列表

在图 2-21 的下方有 4 个区域，分别为“报表筛选”“行标签”“列标签”“数值”，代表了数据透视表的 4 个区域。

第三步：根据数据分析的要求选择报表字段，并拖曳到对应的位置，做出透视表。

可以根据需要及目的灵活选择。各区域的字段可以随意变换或删除，需要就拽进来，不需要就拽出去。各字段后面下拉菜单中的选项也可以随意设置，每一次变化都可以得到不同的数据透视表。

第四步：在透视表的基础上稍做调整，做出频数分布表。

例 2-4　图 2-22 所示是 2012 年上半年通辽市福特汽车销售部分数据的截图。根据数据，请完成以下任务。

任务 1：请汇总每个月份每个子品牌的销售情况，并在此基础上制作每个月份汽车销售情况的频数分布表。

任务 2：请汇总上半年每种品牌车辆的销售情况。

任务 3：请汇总上半年每种颜色车辆的销售情况。

任务 4：请汇总每个子品牌每种颜色车辆的销售情况。

分析：以上 4 个任务通过数据透视表可以轻松完成。下面以任务 1 为例详细介绍。

第一步：按照上述方法，创建一个空白透视表。

	A	B	C	D	E	F	G	H
1	销售月份	子品牌	车型配置	车型代码	颜色	到库日期	库龄天数	销售价格（万元）
2	1月份	嘉年华	三厢1.3L手动风尚型	85744	炫耀银	40674	236	23.86
3	1月份	嘉年华	三厢1.3L手动风尚型	85744	炫耀银	40674	242	22.58
4	1月份	蒙迪欧致胜	GTDI 200 豪华型浅内饰	86532	银砂黑	40735	183	20.47
5	1月份	蒙迪欧致胜	三厢2.3L自动时尚型	86526	银砂黑	40767	152	19.88
6	1月份	蒙迪欧致胜	GTDI 200 时尚型	86557	银砂黑	40887	37	19.78
7	1月份	嘉年华	三厢1.5L自动时尚型	85708	炫耀银	40674	253	19.68
8	2月份	福克斯	两厢1.8L自动时尚型	86L73	激情红	40735	211	19.58
9	2月份	嘉年华	两厢1.5L自动时尚型	85778	烈焰红	40815	140	19.38
10	2月份	福克斯	三厢1.8L自动时尚型	86L68	星光银	40801	156	18.78
11	3月份	蒙迪欧致胜	三厢2.3L自动时尚型	86526	银砂黑	40735	234	17.98
12	3月份	蒙迪欧致胜	GTDI 200 豪华型浅内饰	86532	银砂黑	40735	237	17.70
13	3月份	福克斯	三厢1.8L自动豪华型	86L69	钛晶灰	40619	358	16.19
14	3月份	嘉年华	两厢1.5L自动运动型	85781	耀目黑	40977	1	15.99
15	3月份	蒙迪欧致胜	GTDI 200 豪华型浅内饰	86560	银砂黑	40887	94	15.68
16	3月份	蒙迪欧致胜	GTDI 240 至尊型浅内饰	86538	钛晶灰	40905	82	15.60
17	3月份	嘉年华	两厢1.5L自动运动型	85783	闪耀黄	40977	10	15.19
18	3月份	福克斯	两厢1.8L手动舒适型	86L71	旋舞橙	40905	83	15.00
19	3月份	福克斯	三厢1.8L手动舒适型	86L66	银砂黑	40801	187	14.80
20	3月份	蒙迪欧致胜	豪华型（深内饰）2.3	86555	银砂黑	40989	0	14.60

图 2-22　2012 年上半年通辽市福特汽车销售部分数据截图

第二步：从“字段列表”中把“月份”拖到“行标签”区域，把“子品牌”拖到“列标签”区域，再把“子品牌”拖到“数值”区域，得到如图 2-23 所示数据透视表。

计数项:子品牌	子品牌				
销售月份	福克斯	嘉年华	蒙迪欧致胜	新福克斯	总计
1月份		3	3		6
2月份	2	1			3
3月份	4	4	9		17
4月份	3	6	1		10
5月份	5	6	4	5	20
6月份	2	2	1	4	9
总计	16	22	18	9	65

图 2-23　“月份，子品牌”数据透视表截图

第三步：根据该数据透视表的信息，制作频数分布表，如表 2-14 所示。

表 2-14　每个月份汽车销售量的频数分布表

销售月份	频数/辆	频率/%	销售月份	频数/辆	频率/%
1	6	9	5	20	31
2	3	5	6	9	14
3	17	26	合计	65	100
4	10	15			

总结：数据透视表(见图 2-23)里包含关于 2012 年上半年通辽市福特汽车销售的详细信息。根据透视表及频数分布表 2-14 可知每个月每个子品牌汽车的销售情况和总销售情况，即上半年共销售 65 辆汽车。其中，5 月销售量最多，为 20 辆(福克斯 5 辆、嘉年华 6 辆、蒙迪欧致胜 4 辆、新福克斯 5 辆)，占总销售量的 31%；其次是 3 月，共销售了 17 辆汽车(福克斯 4 辆、嘉年华 4 辆、蒙迪欧致胜 9 辆)，占总销售量的 26%；2 月销售量最少，仅销售了 3 辆汽车，占总销售量的 5%。

注意：数据透视表可根据需求任意调整，通过“选项”卡和“设计”卡的功能实现。另外，单击行标签和列标签后的倒三角按钮，可按照需求排序或显示数据。除此之外，数据透视表

还有很多功能可以挖掘。比如，它可以做双层行标签与双层列标签复杂交叉的信息表，应用于问卷调查数据的汇总等，这里不再详细介绍。

4. 利用 Excel 的“数据分析”工具库制作频数分布表

例 2-5 请根据例 2-4 的数据，针对销售价格数据，以 3 为组距，统计每个价格区间的汽车销售情况，并制作频数分布表及直方图。

分析：解决这个问题的操作步骤如下。

第一步：对“销售价格”排序(降序或升序)，发现价格变动范围是[7.4,23.86]。因此，分成以下六个组(区间)：[7,10]、[10,13]、[13,16]、[16,19]、[19,22]、[22,25]。

第二步：把上述分组输入到 Excel 数据表中，作为接收区间，如图 2-24 所示。

H	I	J
销售价格（万元）		接收区域
23.86		7
22.58		10
20.47		13
19.88		16
19.78		19
19.68		22
19.58		25
19.38		

图 2-24 输入接收区域

注意：在接收区域中输入各组的上限。

第三步：单击“数据分析”，选择“直方图”，如图 2-25 所示，打开“直方图”对话框，输入相应区域进行设置：输入区域 $H$1:$H$66，接收区域 $J$1:$J$8。若包含列标志，就勾选“标志”，输出选项按需求勾选，如图 2-26 所示。

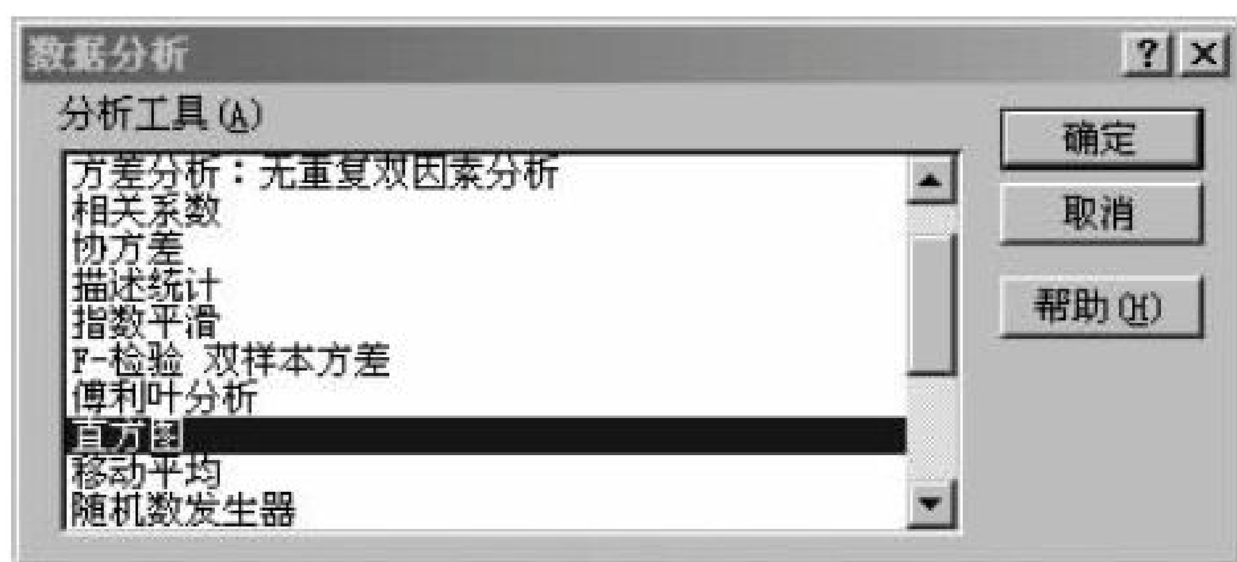


图 2-25 “数据分析”对话框



图 2-26 “直方图”对话框

注意：勾选“累积百分率”，则在频数分布表的结果中将有累积百分比数值，并在直方图中出现累积百分比折线图；勾选“图表输出”，则输出频数分布表的同时，生成直方图。

第四步：单击“确定”按钮，得到组距为 3 的频数统计表和直方图，如图 2-27 和图 2-28 所示。

区域	频率	累积 %
7	0	0.00%
10	25	38.46%
13	17	64.62%
16	11	81.54%
19	4	87.69%
22	6	96.92%
25	2	100.00%
其他	0	100.00%

图 2-27 频数统计表

注意：

(1) 图 2-27 和图 2-28 中的“频率”就是“频数”，“累计%”就是“累计频率”。图 2-27 中的“其他”项没用，可直接删除。

(2) 直方图的直条之间一般间距为零。可用如下方法绘制间距为零的直方图：双击直方图的任一直条，进入“设置数据系列格式”，将分类间距改为“0”，得到无间隔的直方图，如图 2-29 所示。这里已经删除了“其他”项和折线图。

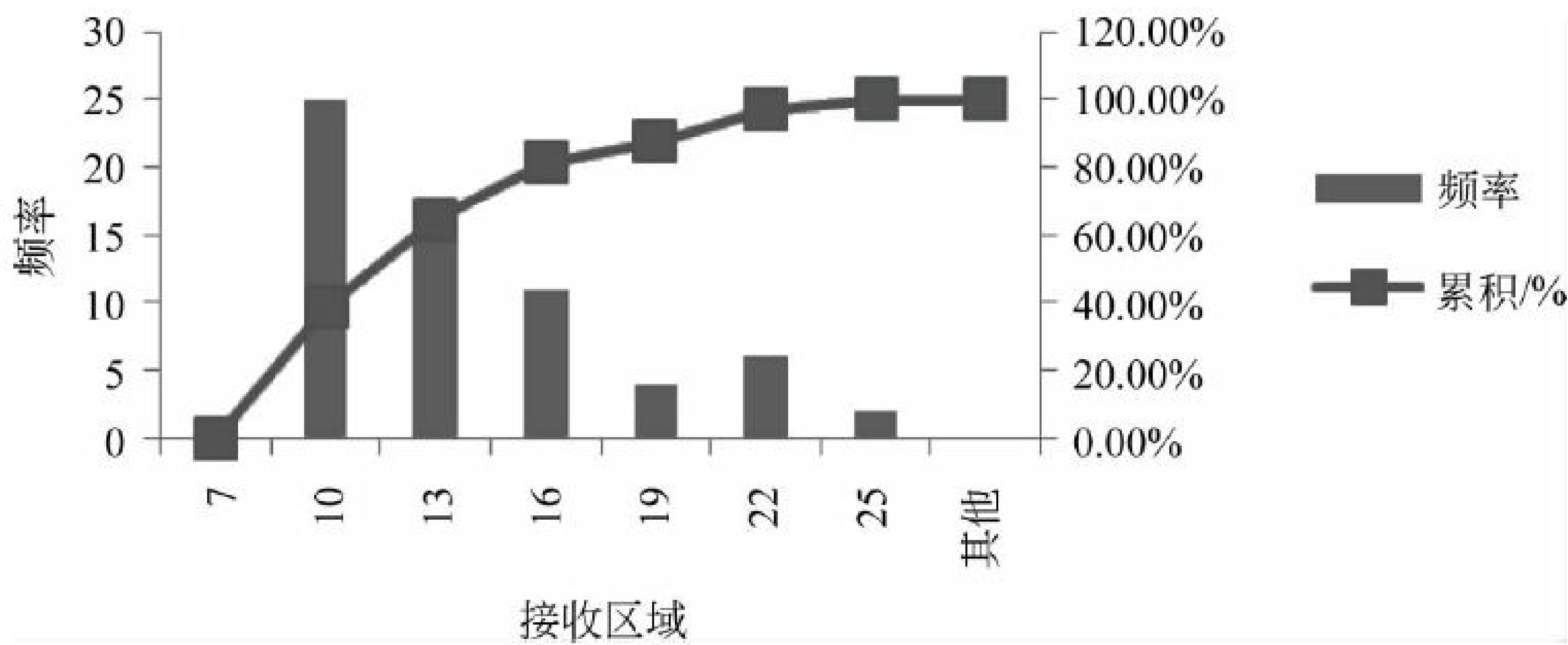


图 2-28　与频数分布表对应的直方图

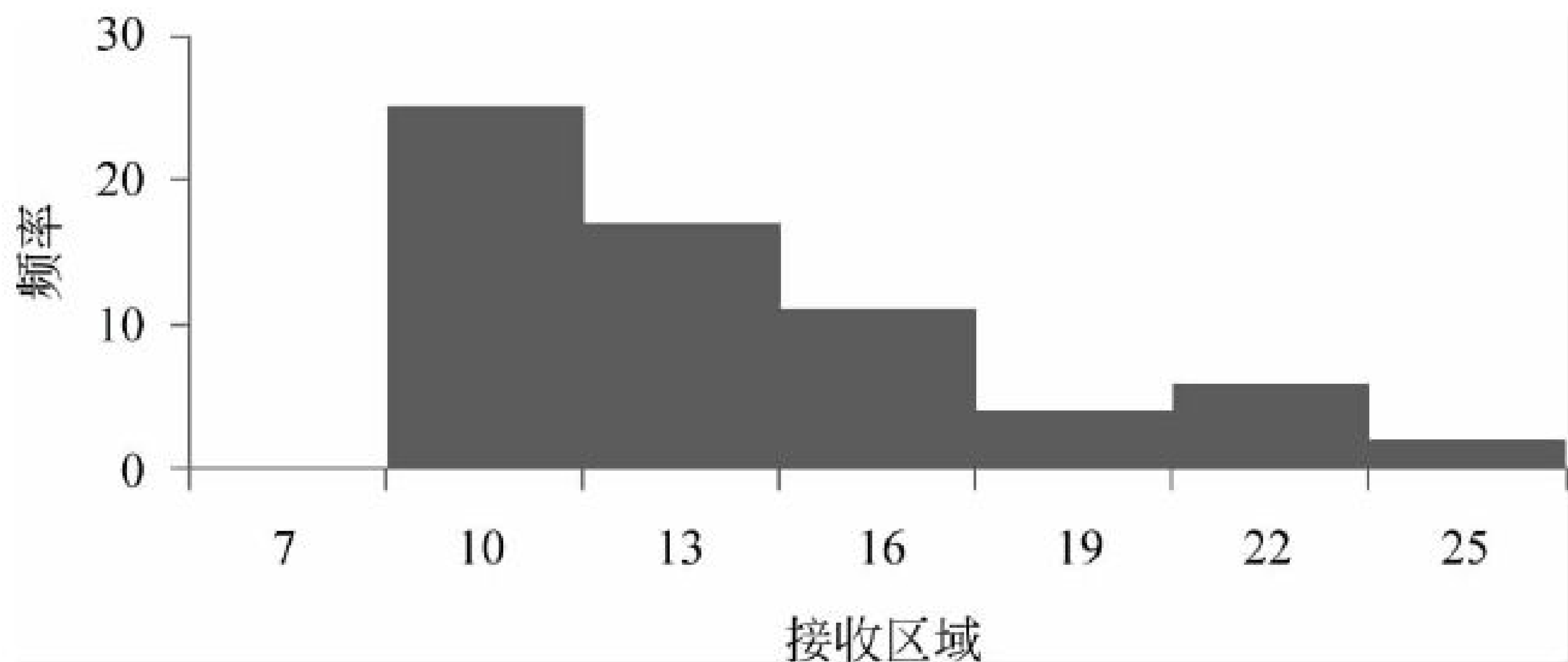


图 2-29　无间隔的直方图

第五步：将图 2-29 中的频数统计表稍做调整，得到如下频数分布表(见表 2-15)。

表 2-15　组距为 3 的频数分布表

销售价格分组/万元	频数/辆	频率/%	销售价格分组/万元	频数/辆	频率/%
7～10	25	38.46	19～22	6	9.23
10～13	17	26.15	22～25	2	3.08
13～16	11	16.92	合　计	65	100
16～19	4	6.15			

总结：从表 2-15 可知，2012 年上半年，在通辽地区共销售了 65 辆福特汽车。其中，价格 7 万～10 万元的汽车销售量最好，共销售了 25 辆，占总销售量的 38.46%；价格 10 万～13 万元的汽车，共销售了 17 辆，占总销售量的 26.15%；价格 13 万～16 万元的汽车销售了 11 辆，占总销售量的 16.92%；价格 22 万元以上的汽车销售量最少，仅为 2 辆，占总销售量的 3.08%。

5. 数据展示图的绘制

前文中介绍了柱形图、条形图、饼图、散点图和折线图等数据展示图，下面介绍如何用 Excel 绘制数据展示图。

各种展示图形的绘制步骤和方法类似，下面以柱形图为例详细介绍。

例 2-6　根据例 2-4 的数据，得到数据透视表，分别为上半年每个月汽车销售数据截图(见图 2-30)和上半年各子品牌明细表截图(见图 2-31)。请根据图 2-30 和图 2-31 所示数据绘制每个月汽车销售量及每个月每个子品牌汽车销售量的柱形图。

	A	B
1	销售月份	频数（辆）
2	1月份	6
3	2月份	3
4	3月份	17
5	4月份	10
6	5月份	20
7	6月份	9
8	合计	65

图 2-30　上半年每个月汽车销售数据截图

	A	B	C	D	E	F
1	计数项:子品牌	子品牌▾				
2	销售月份 ▾	福克斯	嘉年华	蒙迪欧致胜	新福克斯	总计
3	1月份		3	3		6
4	2月份	2	1			3
5	3月份	4	4	9		17
6	4月份	3	6	1		10
7	5月份	5	6	4	5	20
8	6月份	2	2	1	4	9
9	总计	16	22	18	9	65

图 2-31　上半年各子品牌汽车销售量明细表截图

具体操作如下所述。

第一步：选择数据区域＄B＄1：＄B＄7。

第二步：单击“插入”，选择“柱形图”，再选择子图表类型。一般选择第一个，得到如图 2-32 所示的柱形图。

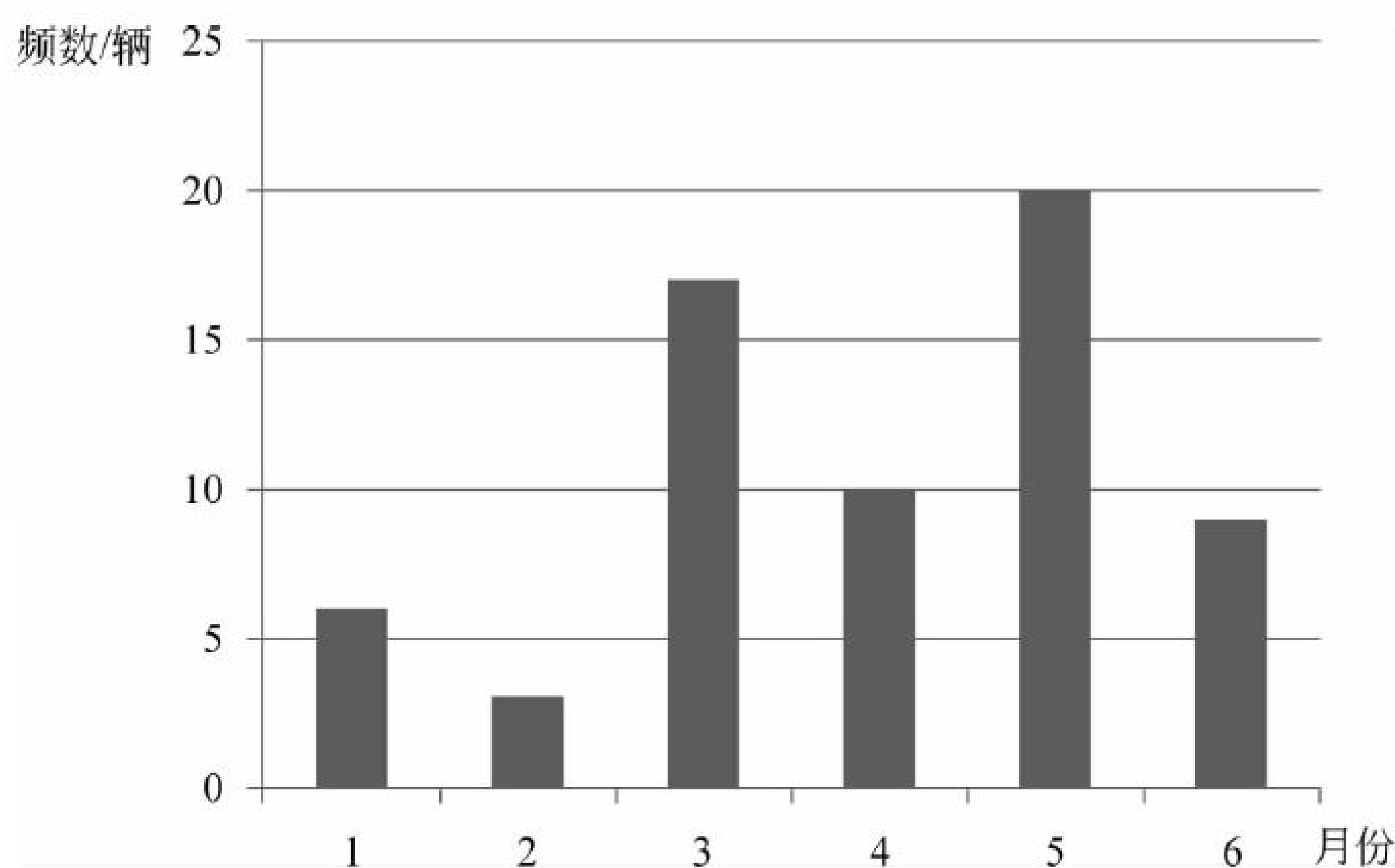


图 2-32　上半年汽车销售量柱形图

类似地，选择数据区域＄B＄2：＄E＄8，单击“插入”，选择“柱形图”，再选择子图表类型，得到如图 2-33 所示的柱形图。

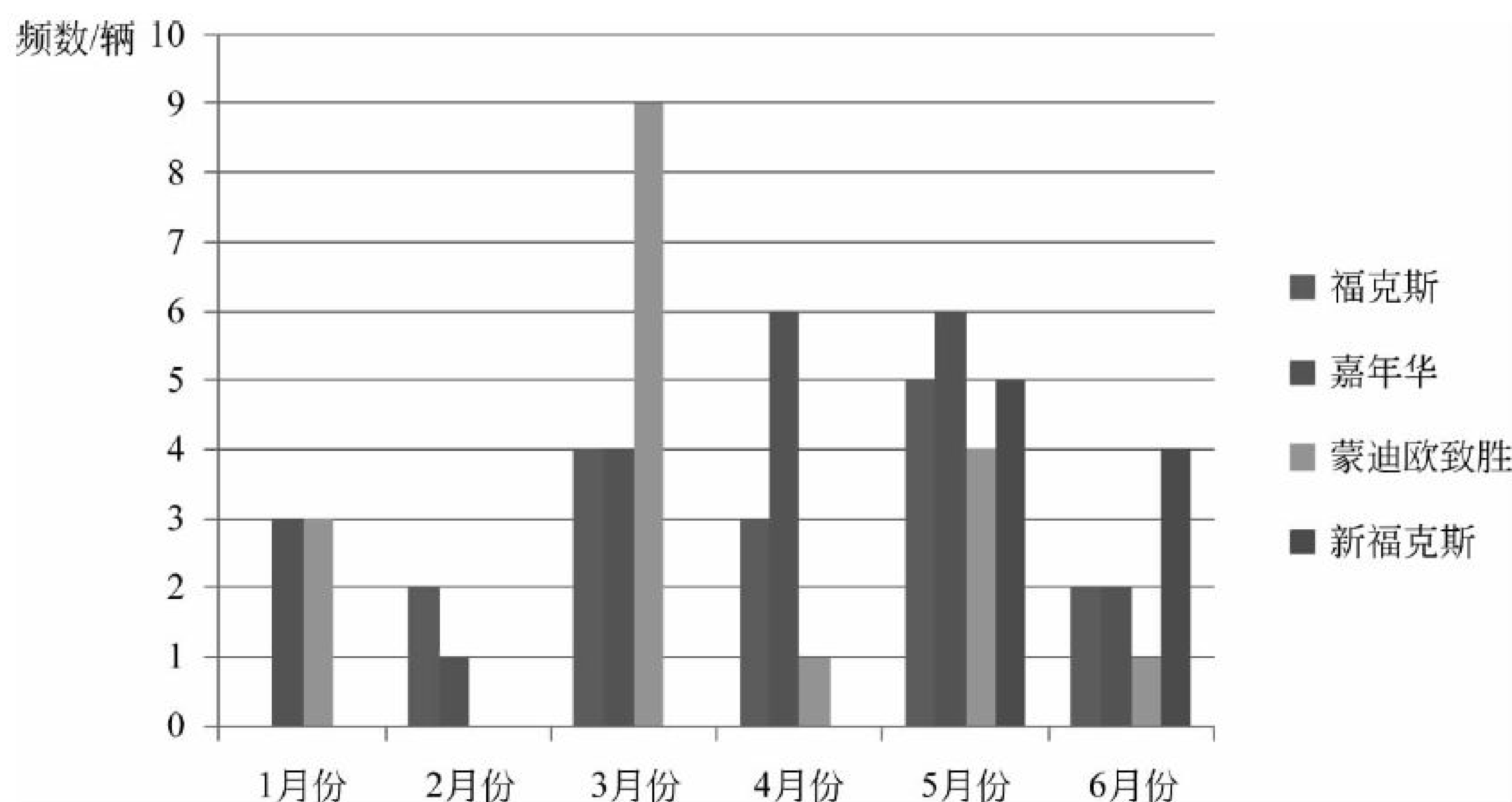


图 2-33　上半年各子品牌汽车销售量明细柱形图

如果对所得到的柱形图的细节之处不满意，比如轴标签、横坐标名称、纵坐标名称等，可再调整。

比如，对于柱形图 2-32，可以把轴标签改为“1 月份，2 月份，…”，也可以加坐标名称，具体方法如下。

第一步：右击图形任一处，选择“选择数据…”，弹出如图 2-34 所示的“选择数据源”对话框。



图 2-34 “选择数据源”对话框

第二步：单击右侧“编辑”，弹出“轴标签”对话框，如图 2-35 所示。在轴标签区域选择“＝Sheet1！$A$2：$A$7”，然后单击“确定”按钮，即可把图 2-32 所示的轴标签改为月份，如图 2-36 所示。

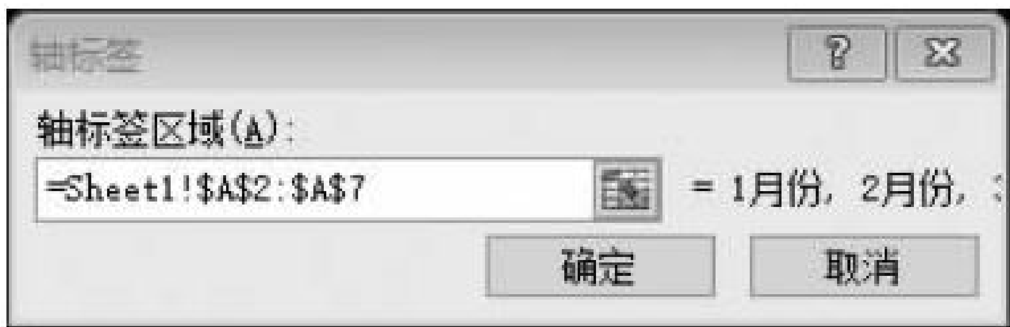


图 2-35 “轴标签”对话框

右击图形纵轴并单击“设置坐标轴格式…”，弹出“设置坐标轴格式”对话框，如图 2-37 所示。可以根据需求，调整纵坐标的各属性。比如，把数字类别改成“数字”，并在小数点后保留两位，得到如图 2-38 所示的柱形图。

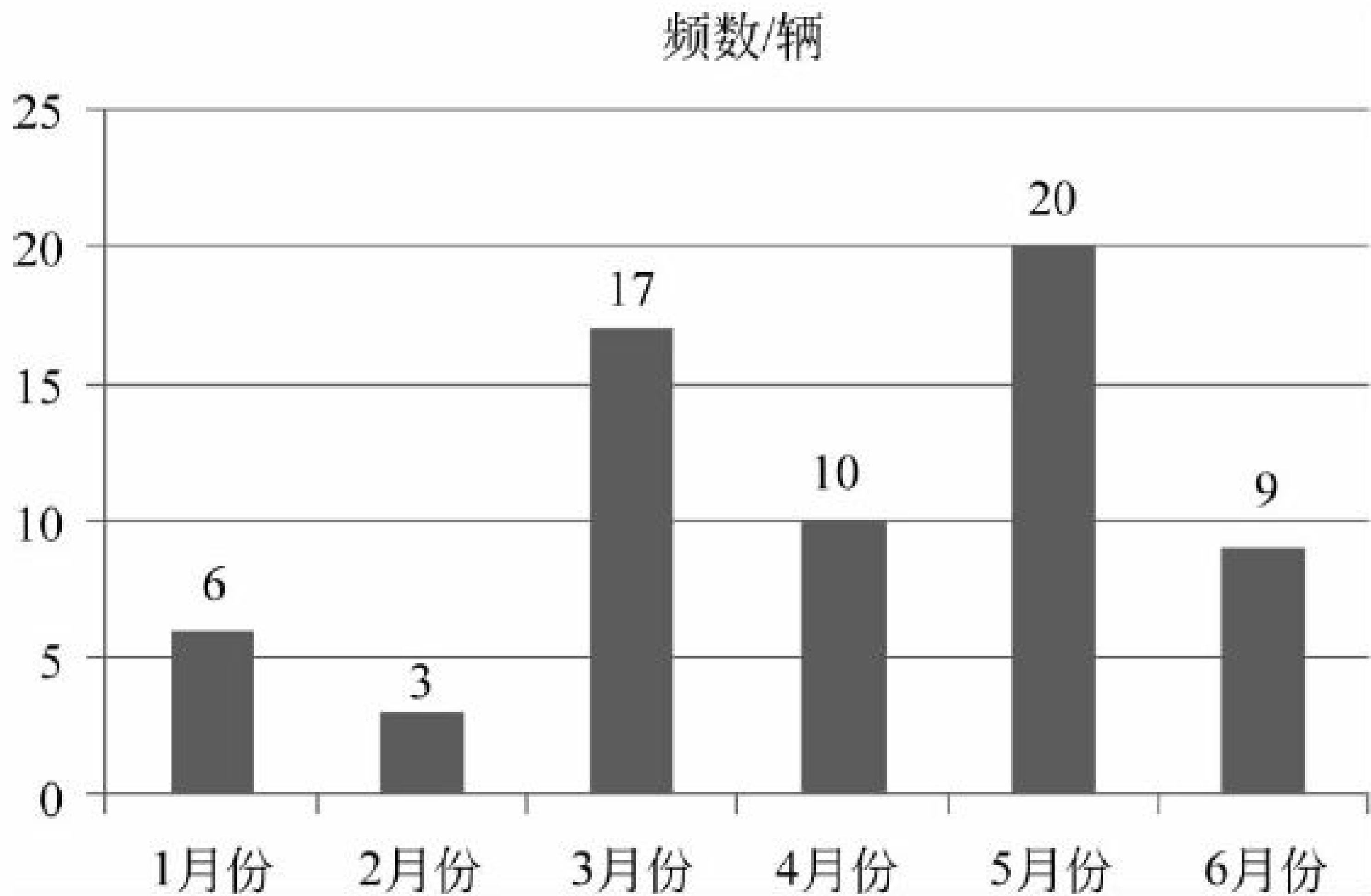


图 2-36 将轴标签改为月份的柱形图

单击图形任一处，调出“设计”“布局”“格式”选项卡，如图 2-39 所示。在此设置图表标题、坐标轴标题等。这里不再详细介绍，请自行完成。



图 2-37 "设置坐标轴格式"对话框

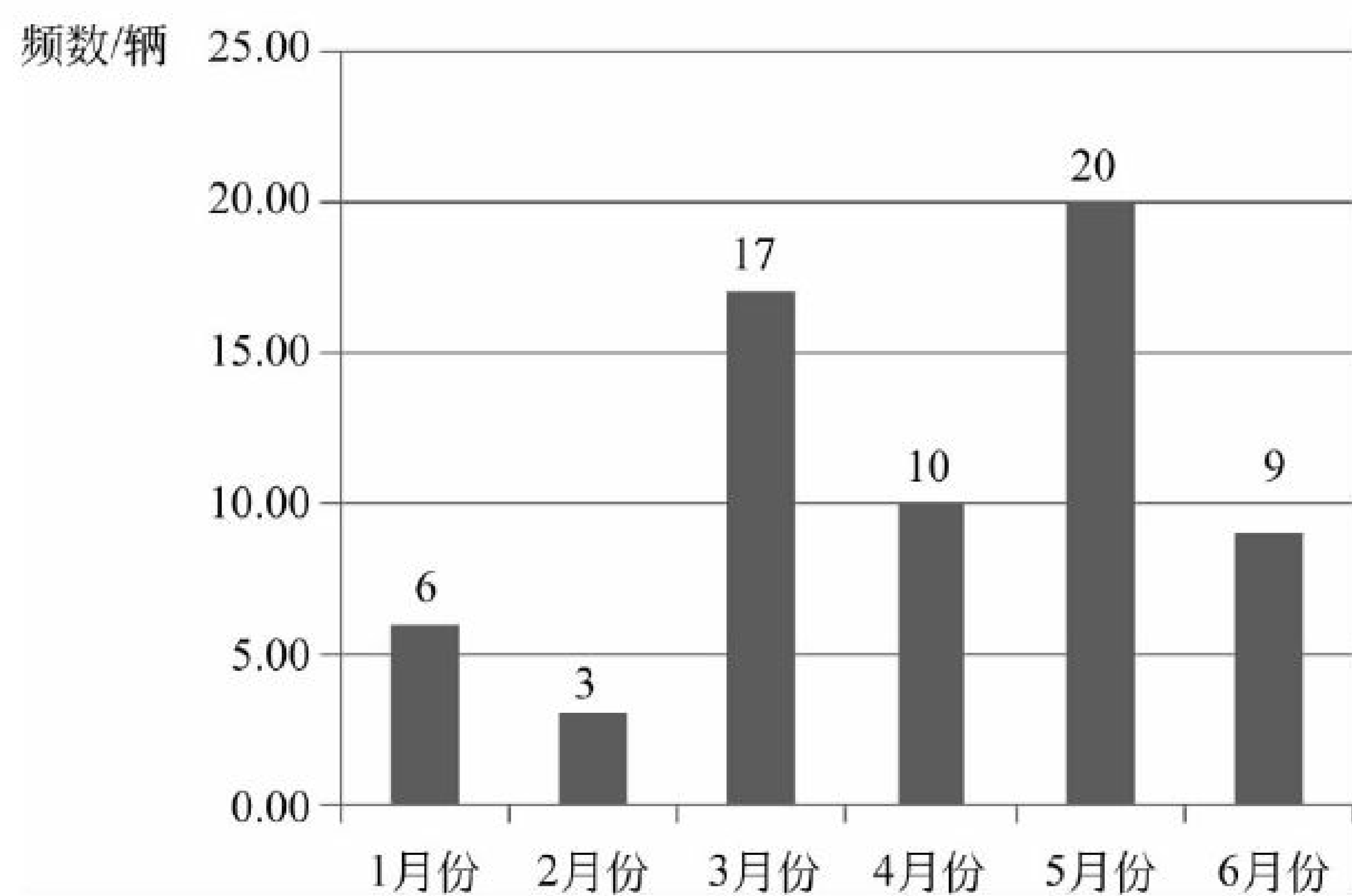


图 2-38 更改了纵坐标数字属性的柱形图



图 2-39 "布局"选项卡

类似上述操作方法和步骤,得到如图 2-40 所示的条形图。

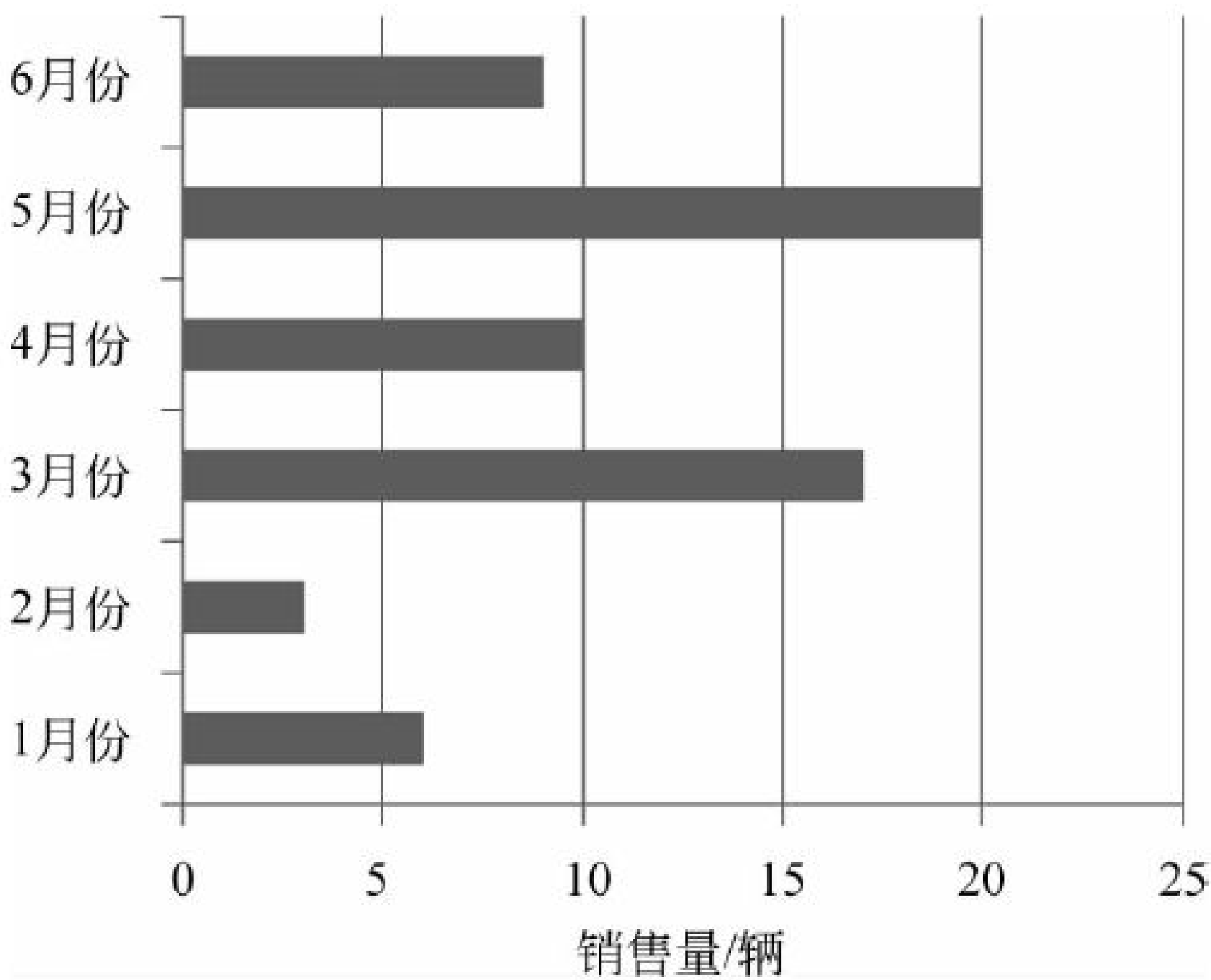


图 2-40　上半年汽车销售量条形图

在图 2-30 中选择数据区域“$A$1:$B$7”,插入“饼图”并右击进行相应的设置,得到如图 2-41 所示的饼图。它形象、直观地展示了每个月汽车销售量占总销售量的比例。

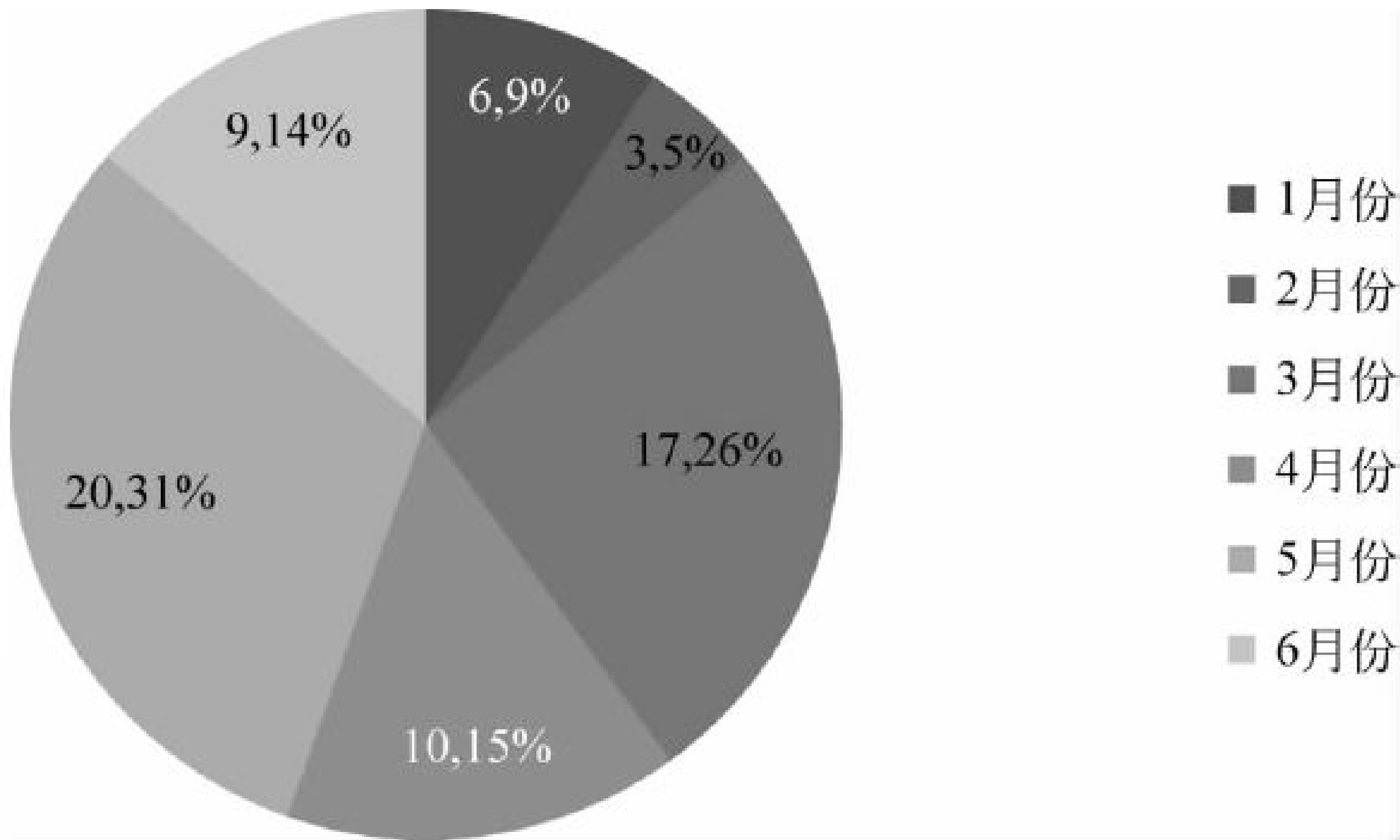


图 2-41　汽车销售量饼图

在图 2-30 中选择数据区域“$A$1:$B$7”,插入“折线图”并右击进行相应的设置,得到如图 2-42 所示的折线图。它形象、直观地展示了汽车销售量随着时间的变化。

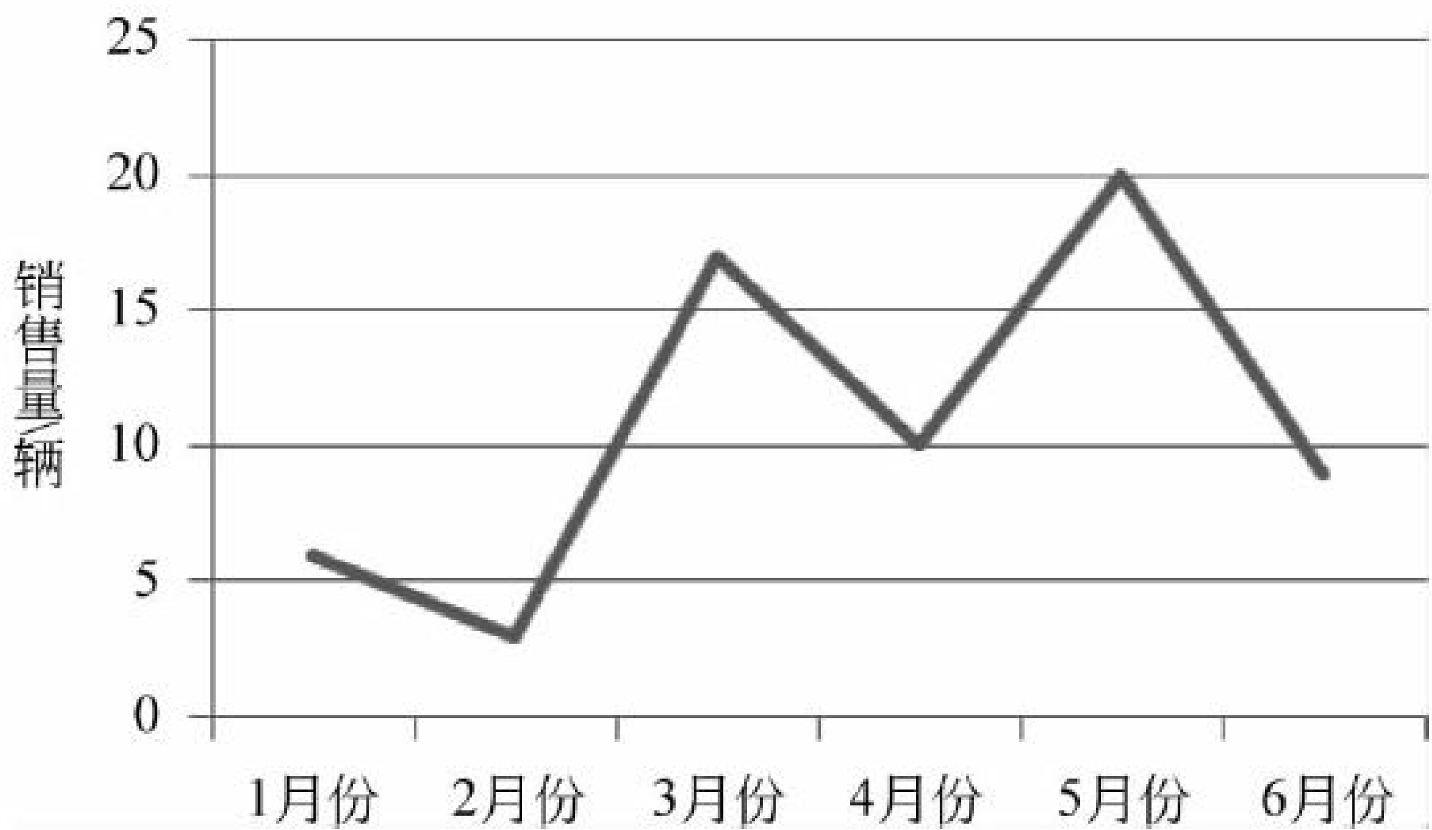


图 2-42　汽车销售量折线图

例 2-7　表 2-16 所示是 2000—2014 年北京市年用电量指标统计数据。请绘制“GDP 与年用电量”“总人口与 GDP”的散点图。

表 2-16　2000—2014 年北京市年用电量指标统计

年份	总人口/万人	GDP/亿元	年用电量/万千瓦小时
2000	1 363.6	3 161.7	3 844 266
2001	1 385.1	3 708.0	3 999 415
2002	1 423.2	4 315.0	4 399 637
2003	1 456.4	5 007.2	4 676 056
2004	1 492.7	6 033.2	5 131 804
2005	1 538.0	6 969.5	5 705 364
2006	1 601.0	8 117.8	6 115 719
2007	1 676.0	9 846.8	6 670 089
2008	1 771.0	11 115.0	6 897 189
2009	1 860.0	12 153.0	7 391 465
2010	1 961.9	14 113.6	8 099 029
2011	2 018.6	16 251.9	8 217 055
2012	2 069.3	17 879.4	8 742 835
2013	2 114.8	19 800.81	9 131 113
2014	2 151.6	21 330.83	9 370 485

资料来源：北京市统计局 2014 统计年鉴，http://www.bjstats.gov.cn/.

选择“GDP”与“年用电量”两列数据，插入“散点图”，并做相应设置，得到如图 2-43 所示的散点图。绘制“总人口与 GDP”散点图的方法类似。

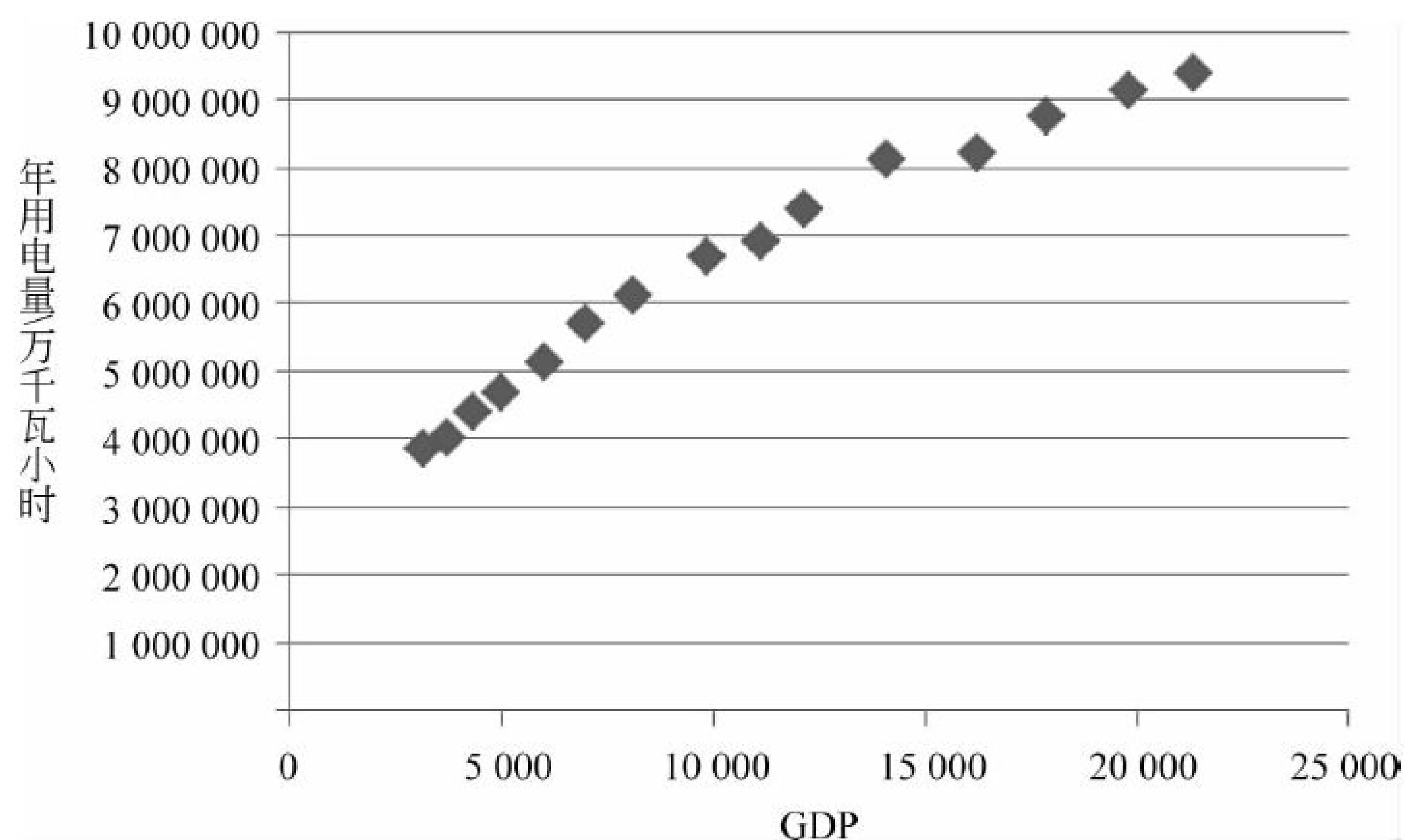


图 2-43　GDP 与年用电量的散点图

6. 利用 Excel 计算常用描述统计量

例 2-8　表 2-1 所示是一个班 46 名学生的高考成绩信息表。请计算出这班学生数学与语文成绩的各描述统计量。

在 Excel 中，利用“数据分析”工具库中的“描述统计”功能（见图 2-44），一次性计算出各

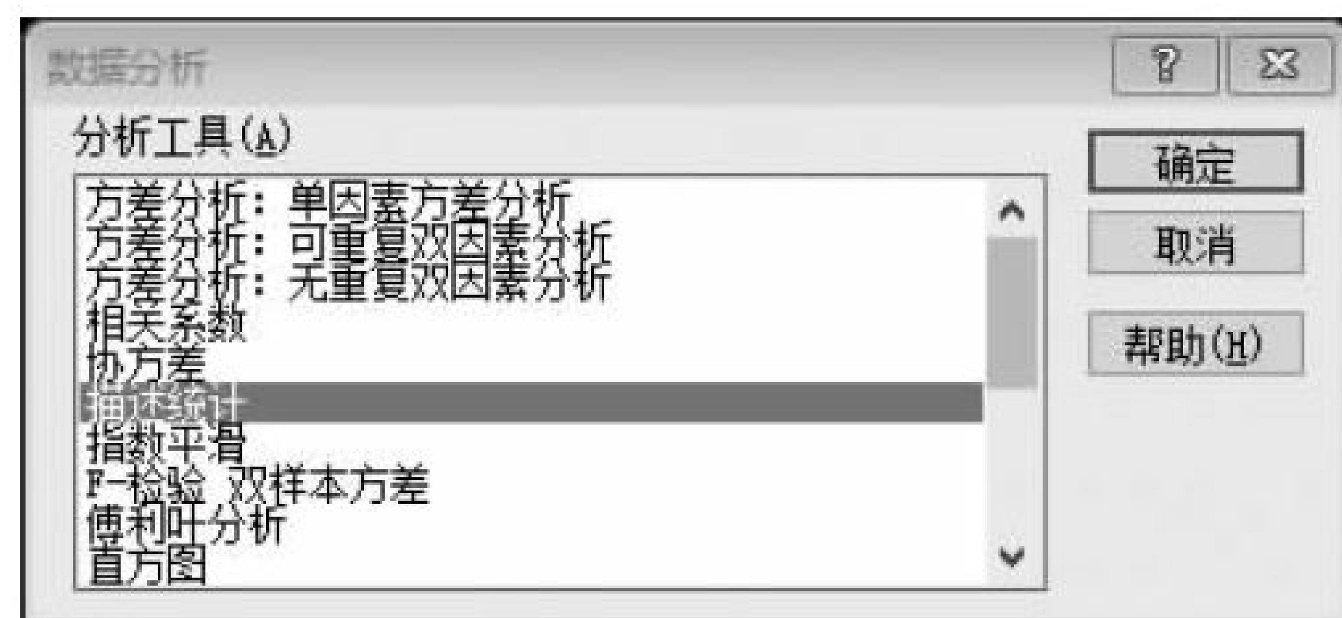


图 2-44 "数据分析"对话框

描述统计量。

在"数据分析"对话框中选择"描述统计"并确定，打开"描述统计"对话框，如图 2-45 所示。在"输入区域"选择"数学"的一列数据，再选择"逐列""标志位于第一行""汇总统计"，"输出选项"可随意三选一，最后单击"确定"按钮，得到如图 2-46 所示各描述统计量。



图 2-45 "描述统计"对话框

	数学	语文
平均	67.69565217	90.63043478
标准误差	2.956180713	1.443077326
中位数	69	93
众数	69	91
标准差	20.04979309	9.787426618
方差	401.9942029	95.79371981
峰度	-0.088376627	1.761901102
偏度	-0.10249439	-1.056619414
区域	93	48
最小值	22	59
最大值	115	107
求和	3114	4169
观测数	46	46

图 2-46 数学和语文成绩的各描述统计量

类似地，得出语文成绩的描述统计量，如图 2-46 第三列所示。

总结：从描述统计结果可知，全班 46 名学生的数学平均成绩为 67.7 分，语文平均成绩为 90.63 分；数学成绩极差为 115－22＝93(分)，语文成绩极差为 107－59＝48(分)；标准差系数分别为 20.05/67.7＝0.30，9.79/90.63＝0.11；对比两门成绩的极差和标准差系数可知，语文成绩的分布更均匀一些，因此语文平均成绩能更好地代表大部分人的语文水平。对于数学而言，用其众数 69 分代表一般水平更合适。峰度和偏度用来描述数据的正态分布情况。更多的详细介绍可参考其他相关书籍。

教学方法和教学策略

教学方法	讲授与探究式教学法
师生角色	学生为主体；教师是启发者和辅导者
实施重点	任务驱动，让学生在实践中探究学习

续表

实施步骤	教学过程与内容	教学方法与策略	学时
1. 准备	(1) 布置学习任务 (2) 引导学生搜集并预习数据处理的相关知识	引导教学法	1
2. 实施	(1) 讲授数据处理的相关理论 (2) 介绍用 Excel 做数据处理的方法 (3) 分发单项任务完成实际操作 (4) 分发综合任务完成实际操作 (5) 综合分析实际操作结果,完成数据处理任务并形成文字解释	讲授法 思维导图 小组展示	5
3. 检查	(1) 对照任务要求,检查数据处理过程与结果 (2) 发现存在的问题与不足,并做出相应改进	问题引导法 对比分析法	1
4. 评价	(1) 师生共同评价任务完成的效果和质量 (2) 针对不足提出改进措施与注意事项	实操展示 小组讨论	1

成果展示要求

展示成果	1. 储药柜的设计(学习任务页-zh2)论文(Word 排版) 2. 储药柜的设计模型建立过程介绍(PPT)	
展示方式	1. 每组把设计方案制作成幻灯片进行展示 2. 数据处理部分进行现场操作展示	
展示顺序	抽签决定	
展示时间	每个小组展示时间 7 分钟,回答问题时间 3 分钟	
展示要求	储药柜设计论文	• 格式正确,排版规范 • 结构完整,有较强的逻辑性和系统性 • 重点突出,包含重要信息和关联点 • 计算准确,论述周详 • 假设合理,推导严密,结论理由充分 • 条理清楚,行文流畅
	PPT 展示	• 格式和视觉效果较好 • 内容完整,条理清楚,逻辑严谨 • 内容、语言和媒体的选择相辅相成 • 体现团队精神 • 语言富有表现力,易于理解 • 有眼神交流,表情生动 • 语速、语调适中

学习评价——个人工作过程报告

学生信息：________________________________

请在下表中认为描述相符的数字位置画上“√”。

1	小组工作中，个人感觉小组氛围如何？						
	非常好	1	2	3	4	5	非常不好
2	小组成员对本人的接受度如何？						
	完全接受	1	2	3	4	5	完全排斥
3	本人的行为是否赢得了小组的理解？						
	非常强	1	2	3	4	5	根本没有
4	在项目中是否学到了东西？						
	非常多	1	2	3	4	5	根本没有
5	对小组工作结果是否满意？						
	非常满意	1	2	3	4	5	完全不满意
6	感觉老师对我们小组工作						
	非常有帮助	1	2	3	4	5	没有帮助、引导

通过下面的调查表，对小组中个人的工作进行评价。

在我的团队中，我能……	评　　价
专业能力 • 按照专业和实际情况，展示学习和工作成果 • 独立或在小组中按照时间要求设计工作步骤 • 充分利用专业信息来源和工作辅助工具 • 想出新创意，给出创造性建议 • 区别重要信息和非重要信息	
方法能力 • 获取、阅读、分析、理解信息资料 • 策划工作步骤，在规定时间内实施 • 根据目标，采用基本的工作技巧 • 组织新观点，对比不同意见和评价 • 采用适当的方法展示自己的观点、想法和结果	
社会能力 • 接受并遵守小组内达成一致的交流和对话准则 • 为自己的观点提出理由和依据 • 能够对不同意见做出适当的反应 • 接受小组分配的任务，并参与任务分配工作 • 正确面对小组内的冲突，并找出解决方法 • 向其他学生提供帮助，向其他学生寻求帮助	

续表

在我的团队中，我能……	评　　价
自我认识能力 • 自己检查工作成果是否合理、是否正确 • 自己认识到当前工作的步骤及缺点 • 了解个人在小组中的位置和贡献 • 自己确立合适的工作和行动目标	

学习评价——小组工作评价

每个小组中应该确定一名小组长，作为此次工作的负责人。每名组员应该在各项任务中轮流担任小组长。小组长要注意，是否考虑到工作中的每一步，并且实施。当小组工作偏离或背离主题时，应该为小组找到正确的方向。

请小组长在下表右侧各列中，自上而下，将已经完成的工作点用“√”的形式表示出来。

学习任务编号		1	2	3	4	5	6	zh1	zh2
计划阶段	分配任务，弄清任务要求								
	确定工作步骤								
	估计所需时间，制定工作时间表								
实施阶段	顺利完成任务								
	互相帮助，并提供咨询								
	认真、仔细地进行和题目相关的工作								
	定时检查工作进度								
	准时准备好报告								
评价阶段	客观地评价工作结果								
	客观地分析合作情况								
	总结对下次小组工作有用的方法								

对自己和小组其他成员的评价：小组中的每个成员都应在小组工作结束后进行此评价。

是否注意到以下小组规则？	学习任务编号（写出评价分数，5＝极好，0＝极差）							
	1	2	3	4	5	6	zh1	zh2
小组顺利地开展工作								
小组成员对工作任务详细讨论，并确定目标								
小组成员互相帮助，并且互相关注								
小组成员听取组员意见并且思考								
小组中的所有成员以同样的工作强度参与工作								
每个组员都记录小组及个人的工作过程								
出现问题时，认真讨论并共同解决								
小组成员专心于主题工作，而没有进行其他活动								
小组讨论中没有出现人身攻击或者谩骂的现象								

第三部分

相关分析

学习情境——中国足球超级联赛成绩分析

2015 年 10 月 31 日中国平安中超联赛结束。图 3-1 所示是 2015 赛季 16 支球队最终战绩统计数据，请分析各项数据与总积分之间的相关性。

名次	球队	场次	胜	平	负	进球	失球	净胜球	积分
1	广州恒大淘宝	30	19	10	1	71	28	43	67
2	上海上港	30	19	8	3	63	35	28	65
3	山东鲁能泰山	30	18	5	7	66	41	25	59
4	北京国安	30	16	8	6	46	26	20	56
5	河南建业	30	12	10	8	35	30	5	46
6	上海绿地申花	30	12	6	12	42	44	-2	42
7	石家庄永昌	30	8	15	7	34	31	3	39
8	重庆力帆	30	9	8	13	37	52	-15	35
9	江苏国信舜天	30	9	8	13	39	48	-9	35
10	长春亚泰	30	8	11	11	39	47	-8	35
11	杭州绿城	30	8	9	13	27	35	-8	33
12	辽宁宏运	30	7	10	13	30	46	-16	31
13	天津泰达权健	30	7	10	13	39	46	-7	31
14	广州富力	30	8	7	15	35	41	-6	31
15	贵州茅台	30	7	8	15	39	52	-13	29
16	上海申鑫	30	4	5	21	30	70	-40	1

图 3-1　2015 中超战绩的统计数据

资料来源：中国平安中超联赛官方网站(http://csl.sina.com.cn/).

能力矩阵——相关分析

学生活动	教师控制/教师组织			自我导向/自我组织		
教师角色	讲授、辅助咨询			教练		
学习步骤	能力水平					
	认知		应用		创新	
	A1	A2	B1	B2	C1	C2
	识记	理解	应用	分析	创造	评价
1. 相关分析的意义、步骤	能记住相关关系的概念、相关分析法适用于解决什么类型的问题，及其解决步骤	能理解相关分析的意义和目的，理解相关关系与函数关系的联系与区别	能判断某实际案例是否可以利用相关分析法解决			
2. 制作散点图	能记住什么叫做散点图	能理解制作散点图的目的	能用 Excel 制作散点图	能根据散点图观察变量之间是否相关，以及相关的类型，比如是否线性相关	若非线性相关，该如何分析	
3. 计算相关系数	能记住相关系数的定义、相关系数的性质	能理解相关系数的公式与计算方法，线性相关判断准则	能用公式或其他方法计算相关系数，比如利用 Excel 软件计算	能根据相关系数对相关关系和所研究的问题进行合理分析，并形成文字		
4. 相关系数的显著性检验	能记住显著性检验的一般步骤、显著性水平的概念及其一般取值	能理解相关系数显著性检验的意义	能应用相关系数检验表确定临界值，并进行显著性检验	能根据决策的结果分析实际问题		

学习步骤计划书

姓　　名		班　　级	
学习日期		汇报日期	
学习任务			

学习目标	
(1) 我的目标是什么？ **注意**：每一个目标都应该以“我能……”开始，并将每一个目标编号；请注明通过哪一个学习任务达到该学习目标。 (2) 我想要学会什么？ **注意**：要给出在哪个时间点达到这个目标。	

学习途径	
(1) 我通过哪些途径获得该目标相关信息？ (2) 我需要什么材料？ (3) 我可以通过哪些方式完成该目标？ (4) 达到此目标的过程中会出现什么障碍？ (5) 在完成此目标的每一天中，我要完成什么？	

学习证明	
(1) 通过何种展示方式，证明我达到这个目标？如论文写作、PPT 展示、流程图绘制等。 (2) 怎样检测我是否达到这个目标？ (3) 请给出预计展示成果的时间。	

学习评价	
(1) 对于自己的学习结果是否满意？ (2) 我通过什么方法评价学习质量？ (3) 我的学习信息是否充足？我如何克服上面提到的障碍？ (4) 我还能做些什么，以完善学习成果？	

学习任务页

学习任务页-1　相关分析的意义、步骤

学生信息________　　任务开始日期________　　任务分值__20 分__

1-A1-1　相关关系指的是什么？类型有哪些？

1-A1-2　相关分析可以解决什么类型的问题？

1-A1-3　相关分析法的步骤、流程是什么？

1-A2-1　相关分析的意义和目的是什么？

1-A2-2　相关关系与函数关系的联系与区别是什么？

1-B1　下列两个问题可否应用相关分析法解决，请分别说明。

（1）某大学随机选取 8 名女大学生，其身高和体重数据如表 3-1 所示。

表 3-1　8 名女大学生的身高和体重数据（判断关系）

编号	1	2	3	4	5	6	7	8
身高/cm	164	165	158	170	174	165	155	171
体重/kg	49	58	50	55	64	60	44	59

根据数据，请分析人的身高和体重是否有关系。有的话，是什么关系？

（2）表 3-2 所示是随机抽取的 8 对母女的身高数据。试根据这些数据，分析母亲的身高和女儿的身高是否有关系。如有，是什么关系？

表 3-2　随机抽取的 8 对母女的身高数据　　单位：cm

编号	1	2	3	4	5	6	7	8
母亲 y	154	157	158	159	161	162	163	164
女儿 x	156	155	159	163	162	165	166	167



学习任务页-2　制作散点图

学生信息__________　　任务开始日期__________　　任务分值＿＿25 分＿＿

2-A1　什么是散点图?

2-A2　制作散点图的目的是什么?

2-B1-1　利用 Excel 软件制作下列问题的散点图。

某大学随机选取 8 名女大学生,其身高和体重数据如表 3-3 所示。

表 3-3　8 名女大学生的身高和体重数据(制作散点图)

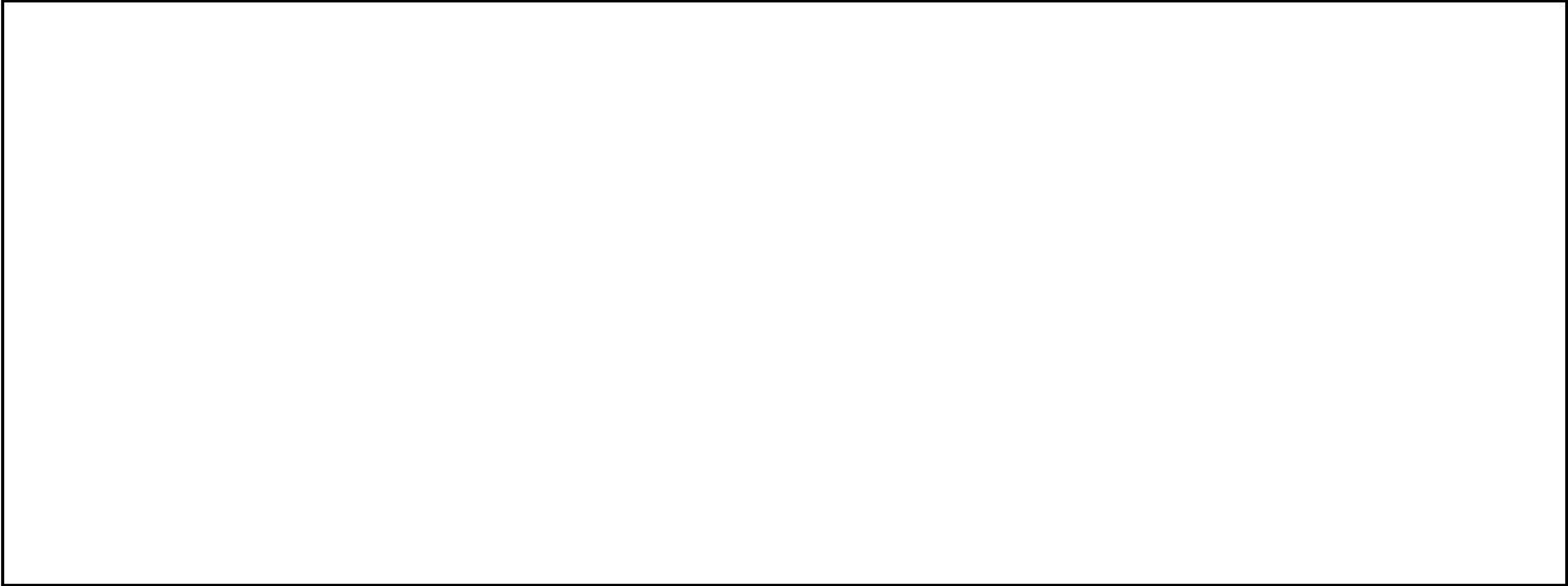
编号	1	2	3	4	5	6	7	8
身高/cm	164	165	158	170	174	165	155	171
体重/kg	49	58	50	55	64	60	44	59

2-B1-2　利用 Excel 软件制作下列问题的散点图。

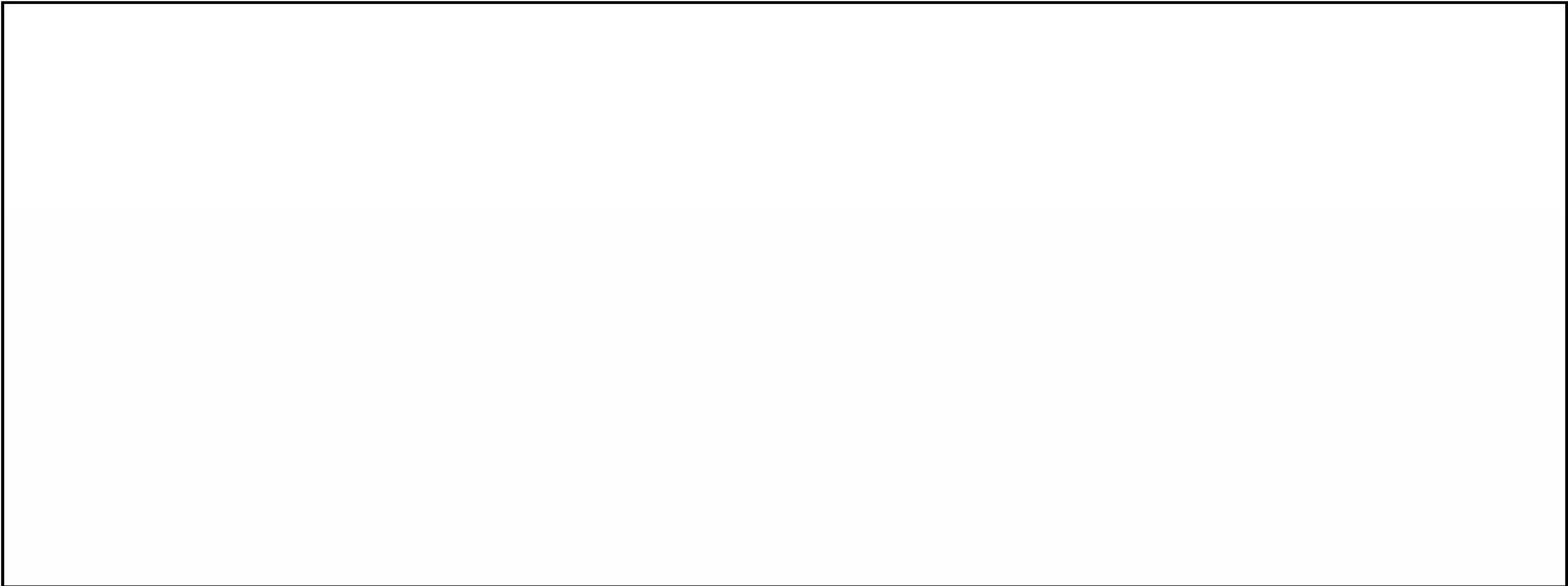
一只红铃虫的产卵数 y 和温度 x 有关，现收集 7 组观测数据列于表 3-4 中。

表 3-4　红铃虫的产卵数和温度数据（制作散点图）

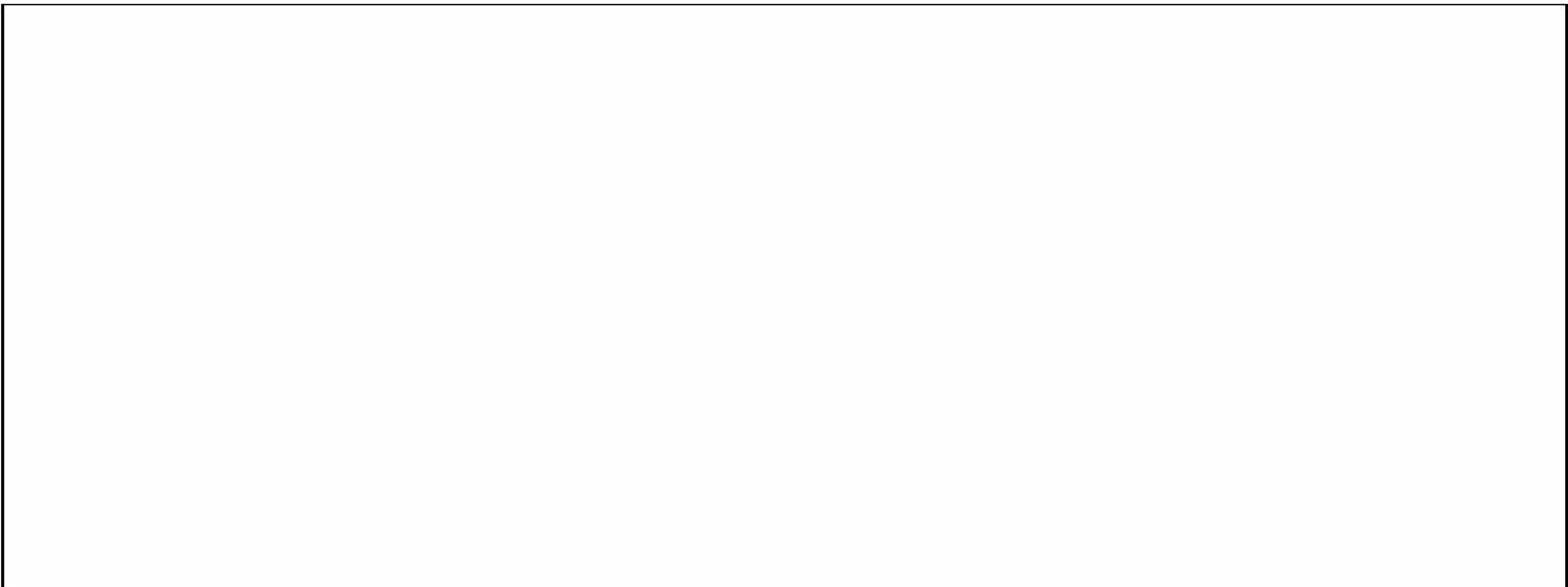
温度 x/℃	21	23	25	27	29	32	35
产卵数 y/个	7	11	21	24	66	115	325



2-B2　根据散点图，初步判断任务“2-B1-1”和“2-B1-2”中的两个问题的相关类型。



2-C1　若非线性相关，该如何处理？



学习任务页-4 计算相关系数

学生姓名________ 任务开始日期________ 任务学时____分钟

★A1 什么是相关系数？它的作用有哪些？

★A1-1 相关系数的计算公式是什么？如何计算？

★A1-2 线性相关的判断准则有哪些？

3-B1-1　请计算下列数据的相关系数(利用 Excel 软件),并判断变量之间相关的密切程度和方向。

某大学随机选取 8 名女大学生,其身高和体重数据如表 3-5 所示。

表 3-5　8 名女大学生的身高和体重数据(计算相关系数)

编号	1	2	3	4	5	6	7	8
身高/cm	164	165	158	170	174	165	155	171
体重/kg	49	58	50	55	64	60	44	59

3-B1-2　请计算下列数据的相关系数(利用 Excel 软件),并判断变量之间相关的密切程度和方向。

一只红铃虫的产卵数 y 和温度 x 有关,现收集 7 组观测数据列于表 3-6 中。

表 3-6　红铃虫的产卵数和温度数据(计算相关系数)

温度 x/℃	21	23	25	27	29	32	35
产卵数 y/个	7	11	21	24	66	115	325

3-B1-3　请计算下列数据的相关系数(利用 Excel 软件),并判断变量之间相关的密切程度和方向。

为了解营业员每人月平均销售额(万元)和利润率(%)之间的关系,从 50 家商店中随机抽取 10 家,得到如表 3-7 所示资料。

表 3-7 10 家店的营业员每人月平均销售额及利润率

商店编号	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
销售额/万元	6	5	8	1	4	8	6	3	4	7
利润率/%	12.5	10.5	18.6	3.1	8.0	16.8	12.2	6.1	7.6	16.9

3-B2-1 请对“任务 3-B1-1”的结果进行合理的文字解释。

3-B2-2 请对“任务 3-B1-2”的结果进行合理的文字解释。

3-B2-3 请对“任务 3-B1-3”的结果进行合理的文字解释。

学习任务页-4　相关系数的显著性检验

学生信息__________　　任务开始日期__________　　任务分值　25 分

4-A1-1　相关系数显著性检验的一般步骤是什么?

4-A1-2　什么叫显著性水平?一般用什么符号表示?取什么值?

4-A2　为什么要做相关系数显著性检验?

4-B1-1　确定“任务 3-B1-1”中的临界值，并进行显著性检验。

4-B1-2　确定“任务 3-B1-2”中的临界值，并进行显著性检验。

4-B2-1　根据“任务 4-B1-1”的显著性检验结果，进一步分析“任务 3-B1-1”中的案例。

4-B2-2　根据“任务 4-B1-2”的显著性检验结果，进一步分析“任务 3-B1-2”中的案例。



学习任务页-zh1　农村居民人均年消费支出和人均年纯收入之间的关系

学生信息__________　　任务开始日期__________　　任务分值___25 分___

表 3-8 所示是 2013 年我国各地区农村居民人均年消费支出和人均纯收入的数据(Excel 数据参见光盘)。请用数学建模的方法研究农村人均年纯收入和人均年消费支出之间的关系。

表 3-8　2013 年我国各地区农村居民人均年消费支出和人均纯收入　　单位：元

地　区	农村居民人均消费支出	农村居民人均纯收入
北　京	13 553.2	18 337.5
天　津	10 155.0	15 841.0
河　北	6 134.1	9 101.9
山　西	5 812.7	7 153.5
内蒙古	7 268.3	8 595.7
辽　宁	7 159.0	10 522.7
吉　林	7 379.7	9 621.2
黑龙江	6 813.6	9 634.1
上　海	142 347	19 595.0
江　苏	9 909.8	13 597.8
浙　江	11 760.2	16 106.0
安　徽	5 724.5	8 097.9
福　建	8 151.2	11 184.2
江　西	5 653.6	8 781.5
山　东	7 392.7	10 619.9
河　南	5 627.7	8 475.3
湖　北	6 279.5	8 867.0
湖　南	6 609.5	8 372.1
广　东	8 343.5	11 669.3
广　西	5 205.6	6 790.9
海　南	5 465.6	8 342.6
重　庆	5 796.4	8 332.0
四　川	6 308.5	7 895.0
贵　州	4 740.2	5 434.0
云　南	4 743.6	6 141.3
西　藏	3 574.0	6 578.2

续表

地　区	农村居民人均消费支出	农村居民人均纯收入
陕　西	5 724.2	6 502.6
甘　肃	4 849.6	5 107.8
青　海	6 060.2	6 196.4
宁　夏	6 489.7	6 931.0
新　疆	6 119.1	7 296.5

资料来源：国家统计局，中国年鉴 2013.

学习任务页-zh2　产品成本与产量关系分析

学生信息__________　　任务开始日期__________　　任务分值　25 分

边际成本是经济学的一个基本概念，它指的是在生产了一定数量的产品后，再生产一个单位的产品时的成本。这说明，产量与成本之间有一定的关系。表 3-9 所示是甲企业产品产量与单位成本数据。请用建模的方法，分析该产品产量与单位成本之间的关系。

表 3-9　甲企业产品产量和单位成本

月数	产量/台	单位成本/(元/台)	月数	产量/台	单位成本/(元/台)
1	4 300	346.23	9	6 024	310.82
2	4 004	343.34	10	6 194	306.83
3	4 300	327.46	11	7 558	305.11
4	5 016	313.27	12	7 381	300.71
5	5 511	310.75	13	6 950	306.84
6	5 648	307.61	14	6 471	303.44
7	5 876	314.56	15	6 354	298.03
8	6 651	305.72	16	8 000	296.21

学习信息库 1——相关分析概述

相关分析是研究变量之间是否有相互关系，以及相互关系密切程度的重要方法。比如，人的体重与身高，学习成绩与学习时间，血压值与年龄、性别、劳动强度、饮食习惯，财政收入与国民收入、工业总产值、农业总产值、总人口、就业人口、固定资产投资等，它们之间是否有相互关系？有的话，密切程度如何？

1. 相关分析

概括地说，相关分析就是通过定量的指标来研究变量之间的相互关系。简单相关分析是研究两个变量间的相互关系，广义相关分析则是研究一个变量与多个变量间的相互关系。

相关分析解决的问题主要包括以下四点：变量间是否存在关系？如存在，是什么关系？关系强度如何？样本反映的变量间的关系能否代表总体变量间的关系？

(1) 函数关系与相关关系

① 函数关系。函数关系是一种严格的依存关系，可以用数学表达式 $y=f(x)$ 来表示，即当 x 确定时，y 有确定的唯一的值与之对应。比如，圆的面积与半径的关系 $s=\pi r^2$，当半径 r 确定时，面积 s 有确定的唯一的值与之对应；商品的收入与销量和价格的关系 $R=Qp$，当销量 Q 与价格 p 确定时，收入 R 有确定的唯一的值与之对应。

② 相关关系。相关关系是一种虽有依存关系，但不完全唯一确定的随机关系。当一个变量或几个变量确定时，与之有相关关系的另一个变量虽然不能完全确定，但在一定范围内与其对应。比如人的身高与体重，一定身高会对应一定范围的体重；人的血压与年龄，一定年龄的人对应一定范围的血压值。

③ 函数关系与相关关系的联系。对具有相关关系的现象进行定量分析时，利用相应函数关系的数学表达式来表明现象之间的定量关系。

相关关系是相关分析的研究对象，函数关系则是相关分析的工具；函数关系是相关关系的特例，即函数关系是完全的相关关系。

④ 函数关系与相关关系的区别。函数关系是指变量之间的关系是确定的；相关关系中的两个变量的关系是不确定的，而是在一定范围内变动。

函数关系变量之间的依存可以用确定的唯一的函数表现出来，可以给定自变量来推算因变量；而相关关系不能完全用确定的唯一的函数表示或推算。

(2) 相关关系的类型

相关关系按不同的标签分成不同的类型，如表 3-10 所示。

表 3-10　相关关系的类型

标签	密切程度	相关方向	相关形式	变量个数
类别	① 完全相关 一种完全相互依存关系，即函数关系 ② 不相关 现象之间没有任何相互依存关系 ③ 不完全相关 现象之间的相互依存关系介于以上两种关系之间，一般的相关现象都是此类关系	① 正相关 变量之间的变动方向一致，类似于单调递增关系 ② 负相关 变量之间的变动方向相反，类似于单调递减关系	① 线性相关 现象之间的关系大致呈现为线性关系，即直线关系，只有变量的一次项 ② 非线性相关 现象之间的关系不是表现为直线关系，而大致表现为曲线关系	① 单相关 两个变量之间的相关 ② 复相关 两个以上变量之间的相关 ③ 偏相关 在复相关中，在固定其余变量的假设下，研究其中两个变量之间的关系

本书主要介绍简单相关关系，即单相关关系。

2. 相关分析的步骤

(1) 制作相关的空间形式图：散点图

散点图用于表示两个变量的数值之间的相关关系。图中的横轴表示一个变量，纵轴表

示另一个变量。

制作散点图的目的是为了直观地观察相关关系，初步判断相关类型。

例 3-1 请根据表 3-11 所示的数据，制作 GDP 和年用电量之间的散点图，消费品零售总额和总人口之间的散点图并初步判断它们之间的相关性。

表 3-11 2000—2014 年北京市年用电量指标统计

年份	消费品零售总额/亿元	GDP/亿元	总人口/万人	年用电量/万千瓦小时
2000	1 658.7	3 161.7	1 363.6	3 844 266
2001	1 831.4	3 708.0	1 385.1	3 999 415
2002	2 005.2	4 315.0	1 423.2	4 399 637
2003	2 296.9	5 007.2	1 456.4	4 676 056
2004	2 626.6	6 033.2	1 492.7	5 131 804
2005	2 911.7	6 969.5	1 538.0	5 705 364
2006	3 295.3	8 117.8	1 601.0	6 115 719
2007	3 835.2	9 846.8	1 676.0	6 670 089
2008	4 645.5	11 115.0	1 771.0	6 897 189
2009	5 387.5	12 153.0	1 860.0	7 391 465
2010	6 340.3	14 113.6	1 961.9	8 099 029
2011	7 222.2	16 251.9	2 018.6	8 217 055
2012	8 123.5	17 879.4	2 069.3	8 742 835
2013	8 872.1	19 800.8	2 114.8	9 131 113
2014	9 638.0	21 330.8	2 151.6	9 370 485

资料来源：北京市统计局 2014 统计年鉴，http://www.bjstats.gov.cn/.

分析：从图 3-2 和图 3-3 可初步判断，GDP 和年用电量之间、消费品零售总额和总人口之间有正线性相关关系。

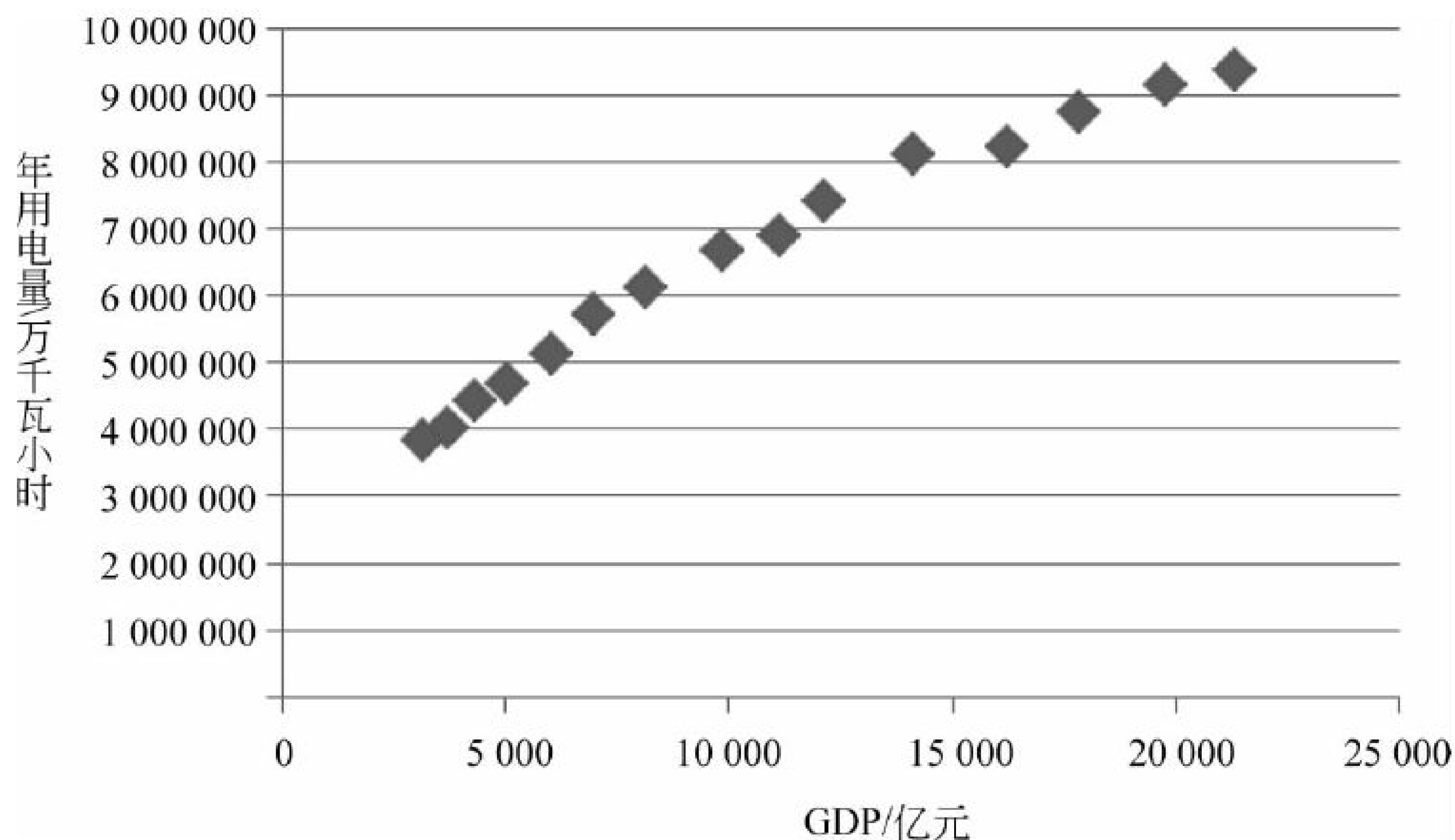


图 3-2 GDP 和年用电量之间的散点图

（2）计算相关系数

简单相关分析是对两个变量间的线性相关程度进行分析，所采用的尺度为单相关系数，简称相关系数。

相关系数是测定两个变量间线性相关关系强度的统计指标。通常用 γ 表示总体的相关

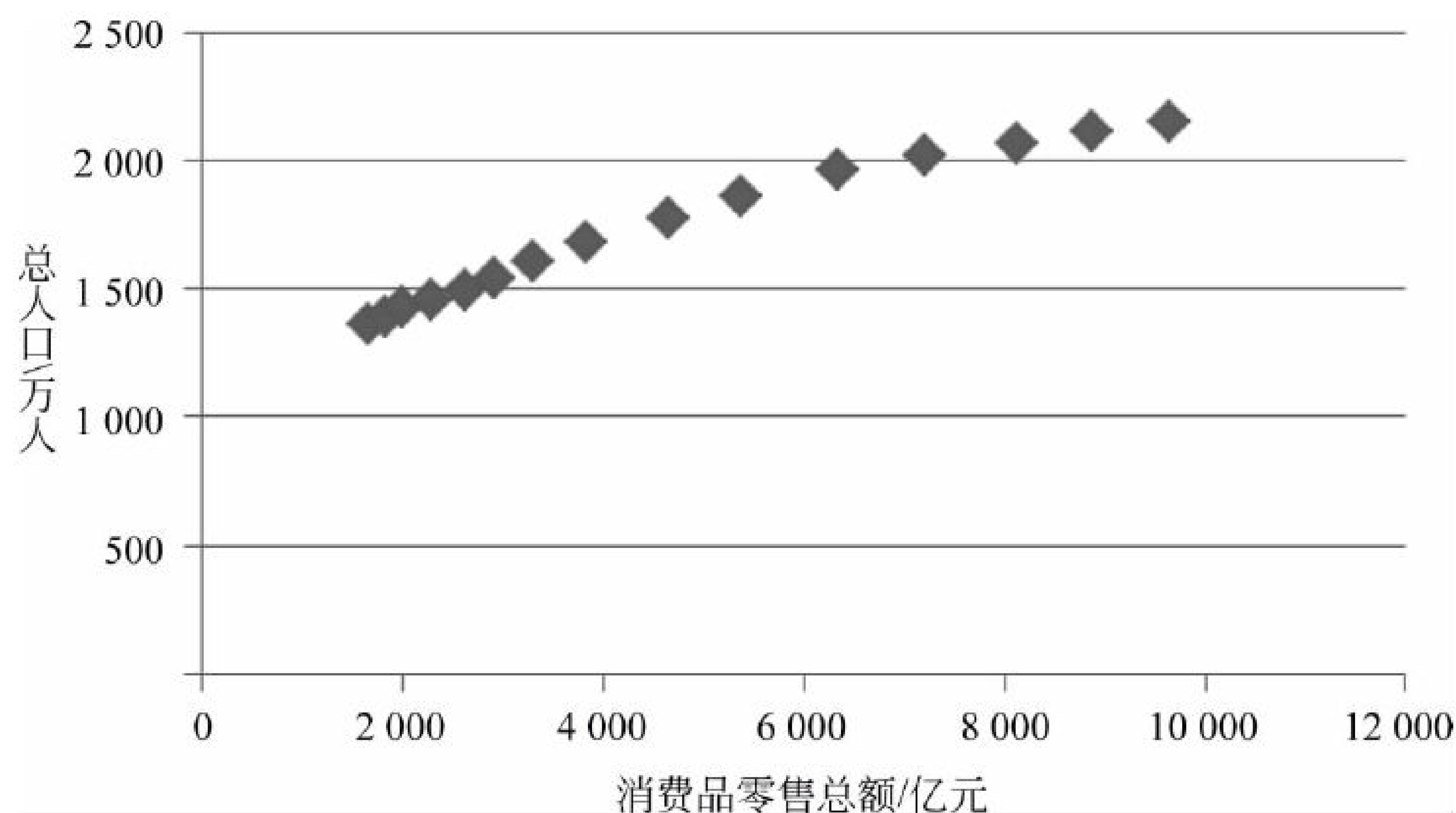


图 3-3　消费品零售总额和总人口之间的散点图

系数，用 r 表示样本的相关系数。一般情况下，讨论样本相关系数。

① 相关系数的计算方法。

定义样本相关系数的公式为

$$r=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2(y_i-\bar{y})^2}} \tag{3-1}$$

其中，$\bar{x}$ 和 $\bar{y}$ 分别是变量 x 和 y 的样本平均数。

通过两个变量与各自平均值的离差的乘积，反映两个变量之间的相关程度，因此，称为积差相关系数。

计算相关系数时，通常利用由式(3-1)推导出来的简化公式，即

$$r=\frac{n\sum x_iy_i-\sum x_i\sum y_i}{\sqrt{\left[n\sum x_i^2-\left(\sum x_i\right)^2\right]\left[n\sum y_i^2-\left(\sum y_i\right)^2\right]}} \tag{3-2}$$

样本相关系数是根据样本观测值计算的，故抽取的样本不同，其具体的数值有所差异。样本相关系数能否代表总体相关系数，需要进一步检验，后面将介绍。

② 相关系数的性质。

a. 相关系数有正负号，分别表示正线性相关和负线性相关。

b. 相关系数 r 的取值范围是[−1，1]，其值大小反映两个变量之间相关关系的密切程度。

c. 相关系数 $|r|=1$ 表明两个变量完全线性相关，此时，线性关系就是函数关系；$r=0$，表明两个变量之间没有线性相关关系，但它不意味两个变量之间不存在其他类型的关系。比如，可能存在非线性关系，需要用其他指标来分析。

③ 线性相关的判断准则。

a. $|r|<0.3$：微弱相关。

b. $0.3\leqslant|r|<0.5$：低度相关。

c. $0.5\leqslant|r|<0.8$：显著相关。

d. $0.8\leqslant|r|<1$：高度相关。

例 3-2　为了解营业员每人月平均销售额(万元)和利润率(%)之间的关系，特从 100 家

商店中随机抽取 10 家，得到如表 3-12 所示数据。试计算样本相关系数。

表 3-12　营业员每人月平均销售额和利润率数据

商店	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
销售额/元	6	5	8	1	4	7	6	3	3	7
利润率/%	12.6	10.4	18.5	3.0	8.1	16.3	12.3	6.2	6.6	16.8

分析：该问题属于计算 10 个样本的样本相关系数的问题，故利用式(3-2)来计算。

解　根据表 3-12 所示数据，列表 3-13。

表 3-13　营业员每人月平均销售额和利润率数据核算

序号	人均销售额 x_i	利润率 y_i	x_i^2	y_i^2	x_iy_i
1	6	12.6	36	158.76	75.6
2	5	10.4	25	108.16	52.0
3	8	18.5	64	342.25	148.0
4	1	3.0	1	9.00	3.0
5	4	8.1	16	65.61	32.4
6	7	16.3	49	265.69	114.1
7	6	12.3	36	151.29	73.8
8	3	6.2	9	38.44	18.6
9	3	6.6	9	43.56	19.8
10	7	16.8	49	282.24	117.6
$\sum$	50	110.8	294	1 465.00	654.9

即

$$\sum x_i = 50,\quad \sum y_i = 110.8,\quad \sum x_i^2 = 294,$$

$$\sum y_i^2 = 1465,\quad \sum x_iy_i = 654.9,\quad n = 10$$

代入式(3-2)，得

$$r=\frac{n\sum x_iy_i-\sum x_i\sum y_i}{\sqrt{\left[n\sum x_i^2-\left(\sum x_i\right)^2\right]\left[n\sum y_i^2-\left(\sum y_i\right)^2\right]}}$$

$$=\frac{10\times 654.9-50\times 110.8}{\sqrt{(10\times 294-50^2)(10\times 1\,465-110.8^2)}}=0.987$$

$$0.8<r=0.987<1$$

因此，根据线性相关的判别准则，初步判断人均销售额与利润率之间存在高度的正相关关系。

注意：这样的计算手动操作比较麻烦，可利用 Excel 软件轻松计算相关系数。

(3) 相关系数的显著性检验

上述例题中的样本相关系数能否代表总体的相关性？样本相关系数带有一定的随机性，样本容量越小，其可信度越差，因此需要检验，称为相关系数的显著性检验。

显著性检验通过了，可根据散点图及相关系数得到变量间的相关关系。但需要注意的是，不是任何两个变量或几个变量之间只要通过了检验，就说明有相关关系，一定要遵循理

论意义和现实经验。比如，人的学习成绩与体重之间的关系或许能够通过相关分析的显著性检验，但从现实经验可知，这两者之间没有任何关系。

相关系数的显著性检验步骤如下所述。

① 计算相关系数 r。

② 确定显著性水平 α，一般 $\alpha=0.05$。

③ 通过相关系数检验临界值表，确定临界值 $r_{\alpha(n-m-1)}$。其中，n 为样本容量，即样本个数；$n-m-1$ 为自由度；m 为自变量的个数。

④ 做出决策。

a. $r \geqslant r_{\alpha(n-m-1)}$：检验通过，表明在给定显著水平 α 下，变量之间的线性关系是显著的。

b. $r < r_{\alpha(n-m-1)}$：检验不通过，表明在给定显著水平 α 下，变量之间的线性关系是不显著的，用样本相关系数代表总体相关系数并不合理。

学习信息库 2——利用 Excel 做相关分析

众所周知，Excel 是常用的办公软件之一，它的普及面广，易于操作。它既是家喻户晓的电子表格软件，也非常适用于基本的统计分析。比如相关分析、回归分析、描述统计等。关于 Excel 在相关分析、回归分析中的应用，本书将简单介绍。

在 Excel 中，可以借助其三个功能做相关分析：一是图表功能；二是函数功能；三是分析工具库。应用图表功能制作散点图。

1. 利用函数计算相关系数

在 Excel 中，可利用 CORREL 函数或 PERSON 函数计算两个变量之间的相关系数。这两个函数是等价的，两个函数的应用方法和结论类似。

应用 CORREL 函数和 PERSON 函数计算相关系数的步骤如下所述。

第一步：依次单击“公式”→“插入函数”→“统计”→“CORREL/PERSON”→“确定”，弹出“CORREL”或“PERSON”对话框。

第二步：在“Array1”和“Array2”中分别输入相应的数据区域并确定，算出相关系数。

例 3-3 图 3-1 所示是 2015 赛季 16 支球队最终战绩统计数据，请计算各项数据之间的相关系数。

利用 CORREL 函数或 PERSON 函数计算两两球队之间的相关系数。用 CORREL 函数计算“负场次”和“积分”之间的相关系数，其他类似。

第一步：在 Excel 工作表中单击任一空白单元格，然后依次单击“公式”→“插入函数”→“统计”→“CORREL”→“确定”，弹出“CORREL”空白对话框，如图 3-4 所示。

第二步：在“Array1”中输入“F1:F17”，在“Array2”中输入“J1:J17”，即可在对话框下方显示出计算结果“−0.952 635 198”，单击“确定”按钮，如图 3-5 所示。

对应的散点图如图 3-6 所示。计算结果 $r=-0.952\ 635\ 198$ 和散点图 3-6 表明，负场次和积分之间负线性相关，即负的场次越多，积分越少。这和直观判断是一致的。这只是初步判断，最终判断需通过相关系数显著性检验来确定。

显著性检验操作如下：



图 3-4　CORREL 空白对话框



图 3-5　CORREL 对话框

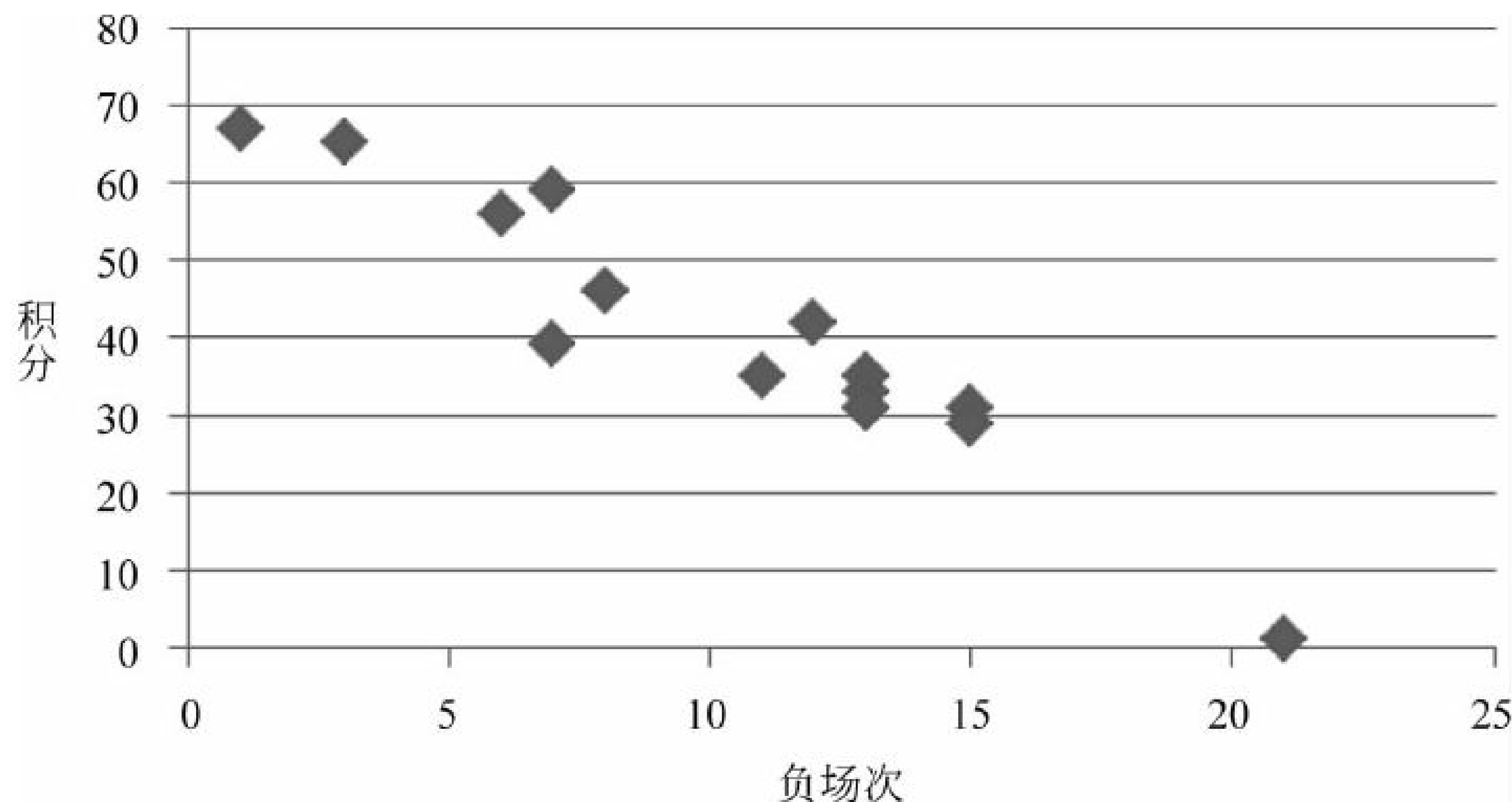


图 3-6　负场次和积分之间的散点图

通过相关系数检验表(见表 3-14)确定临界值 $r_{\alpha(n-m-1)}$。本题中,$n=16$,$m=1$,显著性水平取 $\alpha=0.05$,则临界值为

$$r_{0.05(16-1-1)}=r_{0.05(14)}=0.497\ 31$$

因此,$|r|=0.953>r_{0.05(14)}=0.497\ 31\geqslant r_{\alpha(n-m-1)}$,即相关系数大于临界值,故检验通过,表明在显著水平 $\alpha=0.05$ 下,负场次和积分之间高度负线性相关关系是显著的,说明负的场次越多,积分越少。

2. 利用“数据分析”工具库计算相关系数

Excel 的“数据分析”工具在默认状态下是不显示的,需要手动加载。

表 3-14　相关系数检验临界值表

自由度	显著性水平 α			自由度	显著性水平 α		
$n-m-1$	0.10	0.05	0.01	$n-m-1$	0.10	0.05	0.01
1	0.987 69	0.996 92	0.999 88	39	0.260 48	0.308 13	0.397 82
2	0.900 00	0.950 00	0.990 00	40	0.257 28	0.304 40	0.393 17
3	0.805 38	0.878 34	0.958 74	41	0.254 19	0.300 79	0.388 68
4	0.729 30	0.811 40	0.917 20	42	0.251 21	0.297 32	0.384 34
5	0.669 44	0.754 49	0.874 53	43	0.248 33	0.293 96	0.380 14
6	0.621 49	0.706 73	0.834 34	44	0.245 55	0.290 71	0.376 08
7	0.582 21	0.666 38	0.797 68	45	0.242 86	0.287 56	0.372 14
8	0.549 36	0.631 90	0.764 59	46	0.240 26	0.284 52	0.368 33
9	0.521 40	0.602 07	0.734 79	47	0.237 73	0.281 57	0.364 62
10	0.497 26	0.575 98	0.707 89	48	0.235 29	0.278 71	0.361 03
11	0.476 16	0.552 94	0.683 53	49	0.232 92	0.275 94	0.357 54
12	0.457 50	0.532 41	0.661 38	50	0.230 62	0.273 24	0.354 15
13	0.440 86	0.513 98	0.641 14	51	0.228 39	0.270 63	0.350 86
14	0.425 90	0.497 31	0.622 59	52	0.226 22	0.268 09	0.347 65
15	0.412 36	0.482 15	0.605 51	53	0.224 11	0.265 61	0.344 53
16	0.400 03	0.468 28	0.589 71	54	0.222 06	0.263 21	0.341 50
17	0.388 73	0.455 53	0.575 07	55	0.220 06	0.260 87	0.338 54
18	0.378 34	0.443 76	0.561 44	56	0.218 12	0.258 59	0.335 66
19	0.368 74	0.432 86	0.548 71	57	0.216 23	0.256 37	0.332 84
20	0.359 83	0.422 71	0.536 80	58	0.214 38	0.254 20	0.330 10
21	0.351 53	0.413 25	0.525 62	59	0.212 58	0.252 09	0.327 43
22	0.343 78	0.404 39	0.515 10	60	0.210 83	0.250 03	0.324 82
23	0.336 52	0.396 07	0.505 18	61	0.209 12	0.248 03	0.322 27
24	0.329 70	0.388 24	0.495 81	62	0.207 45	0.246 06	0.319 78
25	0.323 28	0.380 86	0.486 93	63	0.205 82	0.244 15	0.317 35
26	0.317 22	0.373 89	0.478 51	64	0.204 23	0.242 28	0.314 97
27	0.311 49	0.367 28	0.470 51	65	0.202 67	0.240 45	0.312 64
28	0.306 06	0.361 01	0.462 89	66	0.201 15	0.238 66	0.310 36
29	0.300 90	0.355 05	0.455 63	67	0.199 67	0.236 91	0.308 14
30	0.295 99	0.349 37	0.448 70	68	0.198 21	0.235 20	0.305 96
31	0.291 32	0.343 96	0.442 07	69	0.196 79	0.233 52	0.303 82
32	0.286 86	0.338 79	0.435 73	70	0.195 39	0.231 88	0.301 73
33	0.282 59	0.333 84	0.429 65	71	0.194 03	0.230 28	0.299 69
34	0.278 52	0.329 11	0.423 81	72	0.192 69	0.228 71	0.297 68
35	0.274 61	0.324 57	0.418 21	73	0.191 39	0.227 16	0.295 71
36	0.270 86	0.320 22	0.412 82	74	0.190 10	0.225 65	0.293 79
37	0.267 27	0.316 03	0.407 64	75	0.188 85	0.224 17	0.291 89
38	0.263 81	0.312 01	0.402 64	76	0.187 61	0.222 72	0.290 04

续表

自由度	显著性水平 α			自由度	显著性水平 α		
$n-m-1$	0.10	0.05	0.01	$n-m-1$	0.10	0.05	0.01
77	0.186 41	0.221 30	0.288 22	115	0.152 81	0.181 64	0.237 27
78	0.185 22	0.219 90	0.286 43	116	0.152 16	0.180 86	0.236 27
79	0.184 06	0.218 53	0.284 68	117	0.151 51	0.180 10	0.235 28
80	0.182 92	0.217 18	0.282 96	118	0.150 87	0.179 34	0.234 31
81	0.181 80	0.215 86	0.281 27	119	0.150 24	0.178 60	0.233 35
82	0.180 70	0.214 57	0.279 61	120	0.149 62	0.177 86	0.232 40
83	0.179 61	0.213 29	0.277 97	121	0.149 00	0.177 13	0.231 46
84	0.178 55	0.212 04	0.276 37	122	0.148 40	0.176 41	0.230 53
85	0.177 51	0.210 81	0.274 79	123	0.147 80	0.175 70	0.229 61
86	0.176 49	0.209 60	0.273 24	124	0.147 20	0.175 00	0.228 70
87	0.175 48	0.208 41	0.271 72	125	0.146 62	0.174 31	0.227 81
88	0.174 49	0.207 25	0.270 22	126	0.146 04	0.173 62	0.226 92
89	0.173 52	0.206 10	0.268 75	127	0.145 47	0.172 95	0.226 05
90	0.172 56	0.204 97	0.267 30	128	0.144 90	0.172 28	0.225 18
91	0.171 62	0.203 86	0.265 87	129	0.144 34	0.171 61	0.224 33
92	0.170 69	0.202 76	0.264 47	130	0.143 79	0.170 96	0.223 48
93	0.169 78	0.201 69	0.263 08	131	0.143 24	0.170 31	0.222 64
94	0.168 88	0.200 63	0.261 72	132	0.142 70	0.169 67	0.221 82
95	0.168 00	0.199 58	0.260 38	133	0.142 17	0.169 04	0.221 00
96	0.167 13	0.198 56	0.259 06	134	0.141 64	0.168 42	0.220 19
97	0.166 27	0.197 55	0.257 76	135	0.141 12	0.167 80	0.219 39
98	0.165 43	0.196 55	0.256 48	136	0.140 60	0.167 19	0.218 60
99	0.164 60	0.195 57	0.255 22	137	0.140 09	0.166 58	0.217 82
100	0.163 78	0.194 60	0.253 98	138	0.139 59	0.165 98	0.217 04
101	0.162 98	0.193 65	0.252 75	139	0.139 09	0.165 39	0.216 28
102	0.162 18	0.192 71	0.251 55	140	0.138 59	0.164 81	0.215 52
103	0.161 40	0.191 79	0.250 36	141	0.138 10	0.164 23	0.214 77
104	0.160 63	0.190 88	0.249 18	142	0.137 62	0.163 65	0.214 02
105	0.159 87	0.189 98	0.248 02	143	0.137 14	0.163 09	0.213 29
106	0.159 12	0.189 09	0.246 88	144	0.136 66	0.162 52	0.212 56
107	0.158 38	0.188 22	0.245 76	145	0.136 20	0.161 97	0.211 84
108	0.157 65	0.187 36	0.244 65	146	0.135 73	0.161 42	0.211 13
109	0.156 93	0.186 51	0.243 55	147	0.135 27	0.160 87	0.210 42
110	0.156 22	0.185 67	0.242 47	148	0.134 82	0.160 33	0.209 73
111	0.155 52	0.184 84	0.241 40	149	0.134 37	0.159 80	0.209 03
112	0.154 83	0.184 02	0.240 35	150	0.133 92	0.159 27	0.208 35
113	0.154 15	0.183 22	0.239 31	151	0.133 48	0.158 75	0.207 67
114	0.153 48	0.182 42	0.238 28	152	0.133 04	0.158 23	0.207 00

续表

自由度	显著性水平 α			自由度	显著性水平 α		
$n-m-1$	0.10	0.05	0.01	$n-m-1$	0.10	0.05	0.01
153	0.132 61	0.157 72	0.206 33	191	0.118 75	0.141 29	0.185 01
154	0.132 18	0.157 21	0.205 68	192	0.118 44	0.140 92	0.184 54
155	0.131 75	0.156 71	0.205 02	193	0.118 14	0.140 56	0.184 06
156	0.131 33	0.156 21	0.204 38	194	0.117 83	0.140 20	0.183 60
157	0.130 92	0.155 71	0.203 74	195	0.117 53	0.139 84	0.183 13
158	0.130 50	0.155 23	0.203 10	196	0.117 23	0.139 49	0.182 67
159	0.130 09	0.154 74	0.202 47	197	0.116 94	0.139 14	0.182 21
160	0.129 69	0.154 26	0.201 85	198	0.116 64	0.138 79	0.181 76
161	0.129 29	0.153 79	0.201 23	199	0.116 35	0.138 44	0.181 31
162	0.128 89	0.153 31	0.200 62	200	0.116 06	0.138 10	0.180 86
163	0.128 50	0.152 85	0.200 02	201	0.018 23	0.010 91	0.002 88
164	0.128 11	0.152 38	0.199 42	202	0.050 68	0.043 32	0.025 81
165	0.127 72	0.151 93	0.198 82	203	0.068 74	0.066 15	0.051 89
166	0.127 34	0.151 47	0.198 23	204	0.079 15	0.080 69	0.072 53
167	0.126 96	0.151 02	0.197 65	205	0.085 73	0.090 38	0.088 07
168	0.126 58	0.150 58	0.197 07	206	0.090 19	0.097 18	0.099 86
169	0.126 21	0.150 13	0.196 50	207	0.093 37	0.102 17	0.108 98
170	0.125 84	0.149 69	0.195 93	208	0.095 73	0.105 95	0.116 18
171	0.125 47	0.149 26	0.195 36	209	0.097 52	0.108 88	0.121 97
172	0.125 11	0.148 83	0.194 80	210	0.098 91	0.111 20	0.126 70
173	0.124 75	0.148 40	0.194 25	211	0.100 01	0.113 07	0.130 62
174	0.124 39	0.147 98	0.193 70	212	0.100 89	0.114 60	0.133 90
175	0.124 04	0.147 56	0.193 15	213	0.101 60	0.115 86	0.136 67
176	0.123 68	0.147 14	0.192 61	214	0.102 17	0.116 90	0.139 03
177	0.123 34	0.146 73	0.192 08	215	0.102 64	0.117 77	0.141 06
178	0.122 99	0.146 32	0.191 54	216	0.103 02	0.118 50	0.142 81
179	0.122 65	0.145 91	0.191 02	217	0.103 32	0.119 11	0.144 32
180	0.122 31	0.145 51	0.190 49	218	0.103 56	0.119 62	0.145 64
181	0.121 97	0.145 11	0.189 98	219	0.103 76	0.120 06	0.146 79
182	0.121 64	0.144 72	0.189 46	220	0.103 91	0.120 42	0.147 80
183	0.121 31	0.144 32	0.188 95	221	0.104 02	0.120 72	0.148 69
184	0.120 98	0.143 93	0.188 44	222	0.104 10	0.120 97	0.149 46
185	0.120 65	0.143 55	0.187 94	223	0.104 16	0.121 17	0.150 15
186	0.120 33	0.143 16	0.187 44	224	0.104 19	0.121 34	0.150 75
187	0.120 01	0.142 78	0.186 95	225	0.104 20	0.121 47	0.151 27
188	0.119 69	0.142 40	0.186 46	226	0.104 19	0.121 57	0.151 73
189	0.119 38	0.142 03	0.185 97	227	0.104 17	0.121 64	0.152 14
190	0.119 06	0.141 66	0.185 49	228	0.104 13	0.121 69	0.152 49

续表

自由度	显著性水平 α			自由度	显著性水平 α		
$n-m-1$	0.10	0.05	0.01	$n-m-1$	0.10	0.05	0.01
229	0.104 08	0.121 72	0.152 79	267	0.098 98	0.117 06	0.151 05
230	0.104 02	0.121 73	0.153 06	268	0.098 82	0.116 89	0.150 87
231	0.103 95	0.121 73	0.153 28	269	0.098 67	0.116 72	0.150 70
232	0.103 87	0.121 70	0.153 48	270	0.098 51	0.116 55	0.150 52
233	0.103 78	0.121 67	0.153 64	271	0.098 36	0.116 38	0.150 34
234	0.103 68	0.121 62	0.153 77	272	0.098 20	0.116 21	0.150 16
235	0.103 58	0.121 56	0.153 88	273	0.098 04	0.116 04	0.149 97
236	0.103 47	0.121 49	0.153 96	274	0.097 89	0.115 86	0.149 79
237	0.103 36	0.121 41	0.154 03	275	0.097 73	0.115 69	0.149 60
238	0.103 24	0.121 32	0.154 07	276	0.097 58	0.115 52	0.149 41
239	0.103 12	0.121 22	0.154 09	277	0.097 42	0.115 35	0.149 23
240	0.103 00	0.121 12	0.154 10	278	0.097 27	0.115 17	0.149 04
241	0.102 87	0.121 01	0.154 10	279	0.097 11	0.115 00	0.148 85
242	0.102 74	0.120 90	0.154 08	280	0.096 96	0.114 83	0.148 66
243	0.102 60	0.120 78	0.154 04	281	0.096 81	0.114 66	0.148 47
244	0.102 47	0.120 65	0.154 00	282	0.096 65	0.114 49	0.148 27
245	0.102 33	0.120 52	0.153 94	283	0.096 50	0.114 32	0.148 08
246	0.102 19	0.120 39	0.153 88	284	0.096 35	0.114 14	0.147 89
247	0.102 04	0.120 25	0.153 80	285	0.096 20	0.113 97	0.147 69
248	0.101 90	0.120 11	0.153 72	286	0.096 05	0.113 80	0.147 50
249	0.101 75	0.119 97	0.153 62	287	0.095 90	0.113 63	0.147 31
250	0.101 60	0.119 82	0.153 52	288	0.095 75	0.113 46	0.147 11
251	0.101 45	0.119 67	0.153 42	289	0.095 59	0.113 29	0.146 91
252	0.101 30	0.119 52	0.153 30	290	0.095 45	0.113 12	0.146 72
253	0.101 15	0.119 37	0.153 18	291	0.095 30	0.112 95	0.146 52
254	0.101 00	0.119 21	0.153 06	292	0.095 15	0.112 78	0.146 33
255	0.100 85	0.119 05	0.152 92	293	0.095 00	0.112 61	0.146 13
256	0.100 69	0.118 89	0.152 79	294	0.094 85	0.112 45	0.145 94
257	0.100 54	0.118 73	0.152 65	295	0.094 70	0.112 28	0.145 74
258	0.100 38	0.118 57	0.152 50	296	0.094 56	0.112 11	0.145 54
259	0.100 23	0.118 41	0.152 35	297	0.094 41	0.111 94	0.145 35
260	0.100 07	0.118 24	0.152 20	298	0.094 26	0.111 78	0.145 15
261	0.099 92	0.118 08	0.152 04	299	0.094 12	0.111 61	0.144 96
262	0.099 76	0.117 91	0.151 89	300	0.093 97	0.111 44	0.144 76
263	0.099 61	0.117 74	0.151 72	301	0.093 83	0.111 28	0.144 56
264	0.099 45	0.117 57	0.151 56	302	0.093 68	0.111 11	0.144 37
265	0.099 29	0.117 40	0.151 39	303	0.093 54	0.110 95	0.144 17
266	0.099 14	0.117 23	0.151 22	304	0.093 40	0.110 79	0.143 98

续表

自由度	显著性水平 α			自由度	显著性水平 α		
$n-m-1$	0.10	0.05	0.01	$n-m-1$	0.10	0.05	0.01
305	0.093 26	0.110 62	0.143 78	343	0.088 21	0.104 78	0.136 64
306	0.093 11	0.110 46	0.143 59	344	0.088 09	0.104 64	0.136 46
307	0.092 97	0.110 30	0.143 39	345	0.087 97	0.104 49	0.136 28
308	0.092 83	0.110 13	0.143 20	346	0.087 84	0.104 35	0.136 11
309	0.092 69	0.109 97	0.143 00	347	0.087 72	0.104 21	0.135 93
310	0.092 55	0.109 81	0.142 81	348	0.087 60	0.104 07	0.135 75
311	0.092 41	0.109 65	0.142 62	349	0.087 48	0.103 93	0.135 58
312	0.092 27	0.109 49	0.142 42	350	0.087 36	0.103 79	0.135 40
313	0.092 13	0.109 33	0.142 23	351	0.087 24	0.103 65	0.135 23
314	0.092 00	0.109 17	0.142 04	352	0.087 12	0.103 51	0.135 05
315	0.091 86	0.109 01	0.141 85	353	0.087 00	0.103 37	0.134 88
316	0.091 72	0.108 86	0.141 65	354	0.086 88	0.103 23	0.134 70
317	0.091 59	0.108 70	0.141 46	355	0.086 77	0.103 09	0.134 53
318	0.091 45	0.108 54	0.141 27	356	0.086 65	0.102 95	0.134 36
319	0.091 31	0.108 38	0.141 08	357	0.086 53	0.102 82	0.134 19
320	0.091 18	0.108 23	0.140 89	358	0.086 41	0.102 68	0.134 01
321	0.091 04	0.108 07	0.140 70	359	0.086 30	0.102 54	0.133 84
322	0.090 91	0.107 92	0.140 51	360	0.086 18	0.102 41	0.133 67
323	0.090 78	0.107 76	0.140 32	361	0.086 07	0.102 27	0.133 50
324	0.090 64	0.107 61	0.140 13	362	0.085 95	0.102 14	0.133 33
325	0.090 51	0.107 46	0.139 95	363	0.085 84	0.102 00	0.133 16
326	0.090 38	0.107 30	0.139 76	364	0.085 72	0.101 87	0.133 00
327	0.090 25	0.107 15	0.139 57	365	0.085 61	0.101 74	0.132 83
328	0.090 12	0.107 00	0.139 38	366	0.085 49	0.101 60	0.132 66
329	0.089 99	0.106 85	0.139 20	367	0.085 38	0.101 47	0.132 49
330	0.089 86	0.106 70	0.139 01	368	0.085 27	0.101 34	0.132 33
331	0.089 73	0.106 54	0.138 83	369	0.085 16	0.101 21	0.132 16
332	0.089 60	0.106 40	0.138 64	370	0.085 05	0.101 08	0.132 00
333	0.089 47	0.106 25	0.138 46	371	0.084 93	0.100 95	0.131 83
334	0.089 34	0.106 10	0.138 27	372	0.084 82	0.100 82	0.131 67
335	0.089 22	0.105 95	0.138 09	373	0.084 71	0.100 69	0.131 50
336	0.089 09	0.105 80	0.137 91	374	0.084 60	0.100 56	0.131 34
337	0.088 96	0.105 65	0.137 73	375	0.084 49	0.100 43	0.131 18
338	0.088 84	0.105 51	0.137 54	376	0.084 38	0.100 30	0.131 02
339	0.088 71	0.105 36	0.137 36	377	0.084 27	0.100 17	0.130 85
340	0.088 59	0.105 22	0.137 18	378	0.084 17	0.100 05	0.130 69
341	0.088 46	0.105 07	0.137 00	379	0.084 06	0.099 92	0.130 53
342	0.088 34	0.104 93	0.136 82	380	0.083 95	0.099 79	0.130 37

续表

自由度	显著性水平 α			自由度	显著性水平 α		
$n-m-1$	0.10	0.05	0.01	$n-m-1$	0.10	0.05	0.01
381	0.083 84	0.099 67	0.130 21	391	0.082 79	0.098 43	0.128 64
382	0.083 74	0.099 54	0.130 05	392	0.082 69	0.098 31	0.128 49
383	0.083 63	0.099 42	0.129 89	393	0.082 59	0.098 19	0.128 34
384	0.083 52	0.099 29	0.129 74	394	0.082 48	0.098 07	0.128 18
385	0.083 42	0.099 17	0.129 58	395	0.082 38	0.097 95	0.128 03
386	0.083 31	0.099 04	0.129 42	396	0.082 28	0.097 83	0.127 88
387	0.083 21	0.098 92	0.129 27	397	0.082 18	0.097 71	0.127 73
388	0.083 10	0.098 80	0.129 11	398	0.082 08	0.097 59	0.127 58
389	0.083 00	0.098 67	0.128 95	399	0.081 98	0.097 47	0.127 42
390	0.082 89	0.098 55	0.128 80	400	0.081 88	0.097 35	0.127 27

应用分析工具库计算相关系数的步骤如下所述。

例 3-4 根据例 3-3 中的数据,请计算 2015 赛季 16 支球队最终战绩各项数据两两之间的相关系数。

应用分析工具库,可以一次性计算出各项数据两两之间的相关系数,称为相关系数矩阵。具体操作如下所述。

第一步:单击“数据”菜单,选择“数据分析”项,在“数据分析”项中选择“相关系数”,单击“确定”按钮,弹出“相关系数”空白对话框,如图 3-7 和图 3-8 所示。



图 3-7 “数据分析”对话框



图 3-8 “相关系数”空白对话框

第二步:在“输入区域”输入“$D$1:$J$17”,或用鼠标拖曳选中数据;“分组方式”选择“逐列”,再选择“标志位于第一行”。选择输出选项后,单击“确定”按钮,计算出各项数据两两之间的相关系数矩阵,如图 3-9 和图 3-10 所示。



图 3-9 “相关系数”对话框

	胜	平	负	进球	失球	净胜球	积分
胜	1						
平	-0.1514	1					
负	-0.87463	-0.34679	1				
进球	0.899903	-0.19413	-0.75869	1			
失球	-0.64778	-0.45291	0.836797	-0.38909	1		
净胜球	0.939262	0.124232	-0.95218	0.861325	-0.80315	1	
积分	0.945121	0.113828	-0.95264	0.819095	-0.79868	0.970224	1

图 3-10　相关系数矩阵

注意：在"输入区域"栏必须输入连续区域，即变量必须按列连续存放。

从相关系数矩阵可知，负场次和积分之间的相关系数为－0.953，与例 3-3 中的计算结果一致；胜场次和积分之间的相关系数为 0.945，等等。

教学方法和教学策略

教学方法	讲授与探究式教学法		
师生角色	学生为主体；教师是启发者和辅导者		
实施重点	任务驱动，让学生在实践中探究学习		
实施步骤	教学过程与内容	教学方法与策略	学时
1. 准备	(1) 布置学习任务 (2) 引导学生搜集并预习相关分析的相关知识	引导教学法	0.5
2. 实施	(1) 讲授相关分析相关理论 (2) 介绍用 Excel 做相关分析的方法 (3) 分发单项任务完成实际操作 (4) 分发综合任务完成实际操作 (5) 综合分析实际操作结果，完成相关分析任务并形成文字解释	讲授法 思维导图 小组展示	2.5
3. 检查	(1) 对照任务要求，检查相关分析过程与结果 (2) 发现存在问题与不足，并做出相应改进	问题引导法 对比分析法	0.5
4. 评价	(1) 师生共同评价任务完成的效果和质量 (2) 针对不足提出改进措施与注意事项	实操展示 小组讨论	0.5

成果展示要求

展示成果	1. 农村居民人均年消费支出和人均年纯收入(学习任务页-zh1)论文(Word 排版) 2. 费用支出和可支配收入模型建立过程介绍(PPT)
展示方式	1. 每组把分析方案制作成幻灯片进行展示 2. 现场操作展示：散点图的绘制和相关系数的计算等

续表

展示顺序	抽签决定	
展示时间	每个小组展示时间 7 分钟，回答问题时间 3 分钟	
展示要求	农村居民人均年消费支出和人均年纯入论文	• 格式正确，排版规范 • 结构完整，有较强的逻辑性和系统性 • 重点突出，包含重要信息和关联点 • 计算准确，论述周详 • 假设合理，推导严密，结论理由充分 • 条理清楚，行文流畅
	PPT 展示	• 格式和视觉效果较好 • 内容完整，条理清楚，逻辑严谨 • 内容、语言和媒体的选择相辅相成 • 体现团队精神 • 语言富有表现力，易于理解 • 有眼神交流，表情生动 • 语速、语调适中

学习评价——个人工作过程报告

学生信息：________________

请在下表中认为描述相符的数字位置画上“√”。

1	小组工作中，个人感觉小组氛围如何？						
	非常好	1	2	3	4	5	非常不好
2	小组成员对本人的接受度如何？						
	完全接受	1	2	3	4	5	完全排斥
3	本人的行为是否赢得了小组的理解？						
	非常强	1	2	3	4	5	根本没有
4	在项目中是否学到了东西？						
	非常多	1	2	3	4	5	根本没有
5	对小组工作结果是否满意？						
	非常满意	1	2	3	4	5	完全不满意
6	感觉老师对我们小组工作						
	非常有帮助	1	2	3	4	5	没有帮助和引导

通过下面的调查表，对小组中个人的工作进行评价。

在我的团队中，我能……	评　　价
专业能力 • 按照专业和实际情况，展示学习和工作成果 • 独立或在小组中按照时间要求设计工作步骤 • 充分利用专业信息来源和工作辅助工具 • 想出新创意，给出创造性建议 • 区别重要信息和非重要信息	
方法能力 • 获取、阅读、分析、理解信息资料 • 策划工作步骤，在规定时间内实施 • 根据目标，采用基本的工作技巧 • 组织新观点，对比不同意见和评价 • 采用适当的方法展示自己的观点、想法和结果	
社会能力 • 接受并遵守小组内达成一致的交流和对话准则 • 为自己的观点提出理由和依据 • 能够对不同意见做出适当的反应 • 接受小组分配的任务，并参与任务分配工作 • 正确面对小组内的冲突，并找出解决方法 • 向其他学生提供帮助，向其他学生寻求帮助	
自我认识能力 • 自己检查工作成果是否合理，是否正确 • 自己认识到当前工作的步骤及缺点 • 了解个人在小组中的位置和贡献 • 自己确立合适的工作和行动目标	

学习评价——小组工作评价

每个小组中应该确定一名小组长，作为此次工作的负责人。每名组员应该在各项任务中轮流担任小组长。小组长要注意，是否考虑到工作中的每一步，并且实施。当小组工作偏离或背离主题时，应该为小组找到正确的方向。

请小组长在下表右侧各列中，自上而下，将已经完成的工作点用“√”的形式表示出来。

学习任务编号		1	2	3	4	zh1	zh2
计划阶段	分配任务，弄清任务要求						
	确定工作步骤						
	估计所需时间，制作工作时间表						

续表

学习任务编号		1	2	3	4	zh1	zh2
实施阶段	顺利完成任务						
	互相帮助，并提供咨询						
	认真、仔细地进行和题目相关的工作						
	定时检查工作进度						
	准时准备好报告						
评价阶段	客观地评价工作结果						
	客观地分析合作情况						
	总结对下次小组工作有用的方法						

对自己和小组其他成员的评价：小组中的每个成员都应在小组工作结束后进行此评价。

是否注意到以下小组规则？	学习任务编号（写出评价分数，5＝极好，0＝极差）					
	1	2	3	4	zh1	zh2
小组顺利地开展工作						
小组成员对工作任务详细讨论，并确定目标						
小组成员互相帮助，并且互相关注						
小组成员听取组员意见并且思考						
小组中的所有成员以同样的工作强度参与工作						
每个组员都记录小组及个人的工作过程						
出现问题时，认真讨论并共同解决						
小组成员专心于主题工作，而没有进行其他活动						
小组讨论中没有出现人身攻击或者谩骂的现象						

第四部分

回归分析

学习情境——人均可支配收入与人均消费性支出的关系分析

表 4-1 所示是 1995—2014 年北京市城镇居民人均可支配收入与人均消费性支出数据。请利用该数据，研究北京市城镇居民年人均可支配收入与人均消费性支出之间的关系，并估计北京市城镇居民的基础消费水平。

表 4-1 1995—2014 年北京市城镇居民人均可支配收入与人均消费性支出数据 单位：元

年份	城镇居民人均可支配收入	城镇居民人均消费性支出
1995	5 868.4	5 019.8
1996	6 885.5	5 729.5
1997	7 813.1	6 531.8
1998	8 472.0	6 970.8
1999	9 182.8	7 498.5
2000	10 349.7	8 493.5
2001	11 577.8	8 922.7
2002	12 463.9	10 285.8
2003	13 882.6	11 123.8
2004	15 637.8	12 200.4
2005	17 653.0	13 244.2
2006	19 978.0	14 825.0
2007	21 989.0	15 330.0
2008	24 725.0	16 460.0
2009	26 738.0	17 893.0
2010	29 073.0	19 934.0
2011	32 903.0	21 984.0
2012	36 469.0	24 046.0
2013	40 321.0	26 275.0
2014	43 910.0	28 009.0

能力矩阵——回归分析

学生活动	教师控制/教师组织			自我导向/自我组织		
教师角色	讲授、辅助咨询			教练		
学习步骤	能力水平					
	认知		应用		创新	
	A1	A2	B1	B2	C1	C2
	识记	理解	应用	分析	创造	评价
1. 回归分析的意义、种类及步骤	能记住回归分析法适用于解决什么类型的问题，以及解决步骤和种类	能理解回归分析的意义，回归分析的主要目的	能判断某实际问题能否利用回归分析方法解决	能分析不同的回归样本有不同的回归方程，适合不同的回归总体	能分析回归分析与相关分析的联系与区别	
2. 确定自变量和因变量	记住何谓自变量(解释变量)和因变量(预测变量)	能理解如何根据预测目标确定因变量和自变量的方法	能应用前面的知识确定实际问题中的因变量和自变量	能合理分析及说明确定因变量与自变量的依据	能通过市场调查和查阅资料等，找到与预测目标相关的各影响因素	
3. 制作散点图	能记住什么是散点图，什么叫回归直线	能理解制作散点图的目的	能用 Excel 制作散点图，并添加回归直线	能根据散点图观察变量之间是否相关、相关的类型，比如是否线性相关		
4. 确定回归方程	能记住一元线性回归、一元线性回归模型的一般形式、回归截距、回归系数等的定义	能理解回归系数的含义、回归系数和相关系数的关系，能理解最小二乘法的基本思想	能用散点图初步确定回归方程的类型；能应用回归截距和回归系数估计公式计算出参数值，并确定回归方程	能针对回归结果进行合理分析	能够将其他类型的回归问题转化为线性回归问题，能够自学掌握多元回归的相关知识	能够评价、分析不同回归模型的优缺点
5. 回归的显著性检验	能记住什么是 F 统计量、F 显著性水平、t 统计量、P 值等	能理解进行回归的显著性检验的原因及步骤	能用公式计算 F 显著性水平、P 值并进行检验	能根据检验结果分析实际问题	能够自学掌握更多的显著性检验方法	能够对回归结果进行分析，并形成文字解释
6. 计算并确定预测值	能记住回归分析的主要目的、预测的分类及预测的方法	能理解预测值只是估计值，不是实际值；它们之间存在一定的误差	掌握计算预测值的方法，并能应用回归方程计算预测值	能对预测结果进行合理分析，并形成文字解释	能计算预测误差，并分析原因	在点估计的基础上，能够做区间估计

续表

学习步骤	能力水平					
	认　知		应　用		创　新	
	A1	A2	B1	B2	C1	C2
	识　记	理　解	应　用	分　析	创　造	评　价
7. 运用 Excel 进行回归分析	能记住运用 Excel 进行回归分析的方法	能够根据不同的要求，选择不同的方法	能够熟练应用不同方法得出回归方程	能够读懂、分析、应用输出结果，并对实际问题进行预测，且形成文字解释	能够利用 Excel 进行多元回归分析	能够进行多元回归预测

学习步骤计划书

姓　　名		班　　级	
学习日期		汇报日期	
学习任务			

学习目标	
(1) 我的目标是什么? **注意**：每一个目标都应该以"我能……"开始，并将每一个目标编号；请注明通过哪一个学习任务达到该学习目标。 (2) 我想要学会什么? **注意**：要给出在哪个时间点达到这个目标。	
学习途径	
(1) 我通过哪些途径获得该目标的相关信息? (2) 我需要什么材料? (3) 我可以通过哪些方式完成该目标? (4) 达到此目标的过程中会出现什么障碍? (5) 在完成此目标的每一天中，我要完成什么?	
学习证明	
(1) 通过何种展示方式，证明我达到这个目标? 如论文写作、PPT 展示、流程图绘制等。 (2) 怎样检测我是否达到这个目标? (3) 请给出预计展示成果的时间。	
学习评价	
(1) 对于自己的学习结果，我是否满意? (2) 我通过什么方法评价自己的学习质量? (3) 我的学习信息是否充足? 我如何克服上面提到的障碍? (4) 我还能做些什么，以完善学习成果?	

学习任务页

学习任务页-1　回归分析的意义、种类及步骤

学生信息__________　　任务开始日期__________　　任务分值　10 分

1-A1-1　回归分析法可以解决什么类型的问题？

1-A1-2　回归分析的种类有哪些？

1-A1-3　回归分析法的步骤、流程是什么？

1-A2-1　回归分析的意义是什么？

1-A2-2　回归分析的主要目的是什么？

1-B1　对于下列两个问题，可否应用回归分析法解决？请分别说明。

（1）在某大学随机选取 8 名女大学生，其身高和体重数据如表 4-2 所示。

表 4-2　8 名女大学生的身高和体重数据（回归分析法应用判断）

编号	1	2	3	4	5	6	7	8
身高/cm	165	165	157	170	175	165	155	170
体重/kg	48	57	50	54	64	61	43	59

根据数据，请预测身高为 172cm 的女大学生的体重。

（2）表 4-3 所示是随机抽取的 8 对母女的身高数据。试根据这些数据，探讨母亲身高与女儿身高之间的关系。

表 4-3　8 对母女的身高数据（回归分析法应用判断）　　单位：cm

编号	1	2	3	4	5	6	7	8
母亲 y	154	157	158	159	160	161	162	163
女儿 x	155	156	159	162	161	164	165	166

1-B2　能否用女大学生身高和体重关系的回归方程来预测女运动员的相关数据？说明原因。

1-C1　请写出回归分析法与相关分析法的联系与区别。

学习任务页-2　确定自变量和因变量

学生信息__________　　任务开始日期__________　　任务分值＿15 分＿

2-A1　在回归模型中，什么是自变量？什么是因变量？举实例说明。

2-A2　根据什么来确定因变量？用什么方法决定自变量？

2-B1　请确定下列问题中的因变量和自变量，并说明原因。

（1）根据表 4-4 所示数据资料，研究生产费用随产量变化的数据关系。

表 4-4　月产量 x 与生产费用 y 的相关数据（确定因变量和自变量）

序号	月产量 x	生产费用 y	x^2	y^2	xy
1	1.2	62	1.44	3 844	74.4
2	2.0	86	4.00	7 396	172.0
3	3.1	80	9.61	6 400	248.0
4	3.8	110	14.44	12 100	418.0
5	5.0	115	25.00	13 225	575.0
6	6.1	132	37.21	17 424	805.2
7	7.2	135	51.84	18 225	972.0
8	8.0	160	64.00	25 600	1 280.0
$\sum$	36.4	880	207.54	104 214	4 544.6

（2）某企业下半年产品产量与单位成本数据如表 4-5 所示。请预测当产量为 7 000 件时，单位成本为多少元？

表 4-5　某企业下半年产品产量与单位成本数据

月 份	产量/千件	单位成本/元
7	3	72
8	4	71
9	3	73
10	4	69
11	5	68
12	6	70

2-B2　请说明“任务 2-B1”中两个问题的依据。

2-C1　通过市场调查和查阅资料等，寻找与预测目标相关的各影响因素。

随着手机在校园里普及，各大手机厂商都将大学生视为巨大的潜在消费群体，在产品功能定位、价格定位上制定了相应的生产和销售策略。请通过查阅资料，确定手机厂商在功能定位与价格定位上应考虑哪些因素。以苹果手机、三星手机与小米手机为例来分析。


学习任务页-3　制作散点图

学生信息__________　　任务开始日期__________　　任务分值＿15 分＿

3-A1　什么是散点图？什么是回归直线？

3-A2　制作散点图的目的是什么？

3-B1-1　用 Excel 制作“任务 1-B1”中第(1)题的散点图，并添加回归直线。

3-B1-2 用 Excel 制作下列问题的散点图，并添加回归直线。

红铃虫的产卵数 y 和温度 x 有关。现收集 7 组观测数据列于表 4-6 中。

表 4-6 红铃虫的产卵数 y 和温度 x 的观测数据(制作散点图并添加回归直线)

温度 x/℃	21	23	25	27	29	32	35
产卵数 y/个	7	11	21	24	66	115	325

3-B1-3 用 Excel 制作下列问题的散点图，并添加回归直线。

某大学随机选取 8 名女大学生，其身高和体重数据如表 4-7 所示。

表 4-7 8 名女大学生的身高和体重数据(制作散点图并添加回归直线)

编号	1	2	3	4	5	6	7	8
身高/cm	164	165	158	170	174	165	155	171
体重/kg	49	58	50	55	64	60	44	59

3-B2 根据散点图，确定以上三个问题的回归类型。



学习任务页-4　确定回归方程

学生信息__________　　任务开始日期__________　　任务分值____20分____

4-A1-1　什么叫一元线性回归？

4-A1-2　一元线性回归模型的一般形式是什么？

4-A1-3　什么叫回归截距、回归系数？它们的计算公式是什么？

4-A2-1　回归系数的含义是什么？经济含义是什么？

4-A2-2　回归系数和相关系数的关系是什么？

4-A2-3　最小二乘法的基本思想是什么?

4-B1-1　根据“任务 3-B1-1”中的散点图,确定其回归模型的类型。

4-B1-2　根据“任务 3-B1-2”中的散点图,确定其回归模型的类型。

4-B1-3　请确定“任务 3-B1-1”中的回归方程。

4-B1-4　请确定“任务 3-B1-3”中的回归方程。

4-B2-1　能利用一元线性回归法进行回归分析的前提是什么？

4-B2-2　根据回归系数的含义，解释“任务 4-B1-3”和“任务 4-B1-4”的回归结果。

4-C1-1　将“任务 4-B1-2”中的模型转化为一元线性回归模型，并确定回归方程。

4-C1-2　什么叫多元线性回归？

4-C2　请写出不同回归模型的优缺点，如线性回归模型、非线性回归等。

学习任务页-5　回归的显著性检验

学生信息__________　　任务开始日期__________　　任务分值　15 分　

5-A1-1　什么是 F 检验、F 统计量和 F 显著性水平?

5-A1-2　什么是 t 检验、t 统计量和 P 值? 它们的计算公式是什么?

5-A1-3　F 检验的步骤有哪些?

5-A1-4　P 值检验的步骤有哪些?

5-A2　为什么要进行回归的显著性检验? 从哪两个方面来检验?

5-B1-1　请计算“任务 4-B1-3”中的 F 显著性水平及 P 值。

5-B1-2　请计算“任务 4-B1-4”中的 F 显著性水平及 P 值。

5-B2-1　请根据“任务 5-B1-1”的结果，对“任务 4-B1-3”中的回归进行显著性检验并分析。

5-B2-2　请根据“任务 5-B1-2”的结果，对“任务 4-B1-4”中的回归进行显著性检验并分析。

5-B2-3　居民住宅区火灾造成的经济损失与该住户到最近的消防中心的距离之间有一定的关系。表 4-8 中列出了 15 个样本点的数据。请根据数据，确定火灾经济损失与最近的消防中心的距离之间的关系。

表 4-8　15 个样本点的火灾经济损失数据

距消防中心距离 x/千米	3.4	1.8	4.6	2.3	3.1	5.5	0.7	3.0
火灾经济损失 y/千元	26.2	17.8	31.3	23.1	27.5	36.0	14.1	22.3
距消防中心距离 x/千米	2.6	4.3	2.1	1.1	6.1	4.8	3.8	
火灾经济损失 y/千元	19.6	31.3	24.0	17.3	43.2	36.0	26.1	

5-C1　除了 F 检验和 P 值检验，还有哪些显著性检验方法？

5-C2-1　请对“任务 5-B2-1”和“任务 5-B2-2”的结果进行合理的解释并形成文字。

5-C2-2　请对“任务 5-B2-3”的结果进行合理的解释并形成文字。



学习任务页-6　计算并确定预测值

学生信息__________　　任务开始日期__________　　任务分值___15 分___

6-A1-1　回归分析的主要目的是什么？

6-A1-2　如何根据回归方程计算预测值？

6-A2　预测值就是实际值吗？

6-B1　请完成下面两个预测问题。

(1) 某大学随机选取 8 名女大学生，其身高和体重数据如表 4-9 所示。

表 4-9　8 名女大学生的身高和体重数据（计算预测值）

编号	1	2	3	4	5	6	7	8
身高/cm	165	165	157	170	175	165	155	170
体重/kg	48	57	50	54	64	61	43	59

根据数据，请预测身高为 172cm 的女大学生的体重。

（2）表 4-10 所示是随机抽取的 8 对母女的身高数据。试根据数据，预测母亲身高为 170cm 时，女儿身高为多少。

表 4-10　8 对母女的身高数据(计算预测值)　　单位：cm

编号	1	2	3	4	5	6	7	8
母亲 y	154	157	158	159	160	161	162	163
女儿 x	155	156	159	162	161	164	165	166

6-B2-1　请对“任务 6-B1”中第(1)题的预测结果进行合理分析，并形成文字解释。

6-B2-2　请对“任务 6-B1”中第(2)题的预测结果进行合理分析，并形成文字解释。

6-C1-1　请试估计“任务 6-B2-1”中的预测误差，并分析原因。

6-C1-2　请试估计“任务 6-B2-2”中的预测误差，并分析原因。

6-C2-1　什么是点估计？什么是区间估计？

6-C2-2　请试对“任务 6-B1”的两个问题进行区间估计。

学习任务页-7 利用 Excel 进行回归分析

学生信息__________ 任务开始日期__________ 任务分值___10 分___

7-A1-1 利用 Excel 做回归分析的方法有哪几种？

7-A1-2 如何利用 Excel 做简单回归分析？

7-A1-3 在 Excel 中，截距函数和斜率函数分别是什么函数？

7-A2 请阐述利用 Excel 做回归分析的三种方法的优缺点。

7-B1-1　请利用表 4-11 中所示数据，以 GDP 为自变量，年用电量为因变量，制作散点图，添加趋势线，写出一元线性回归方程。

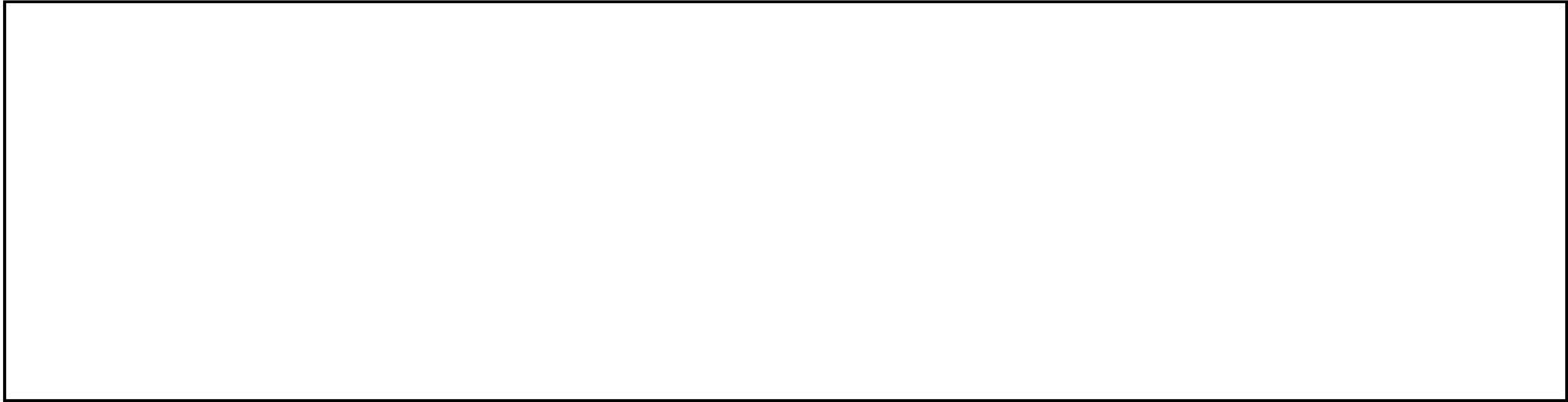
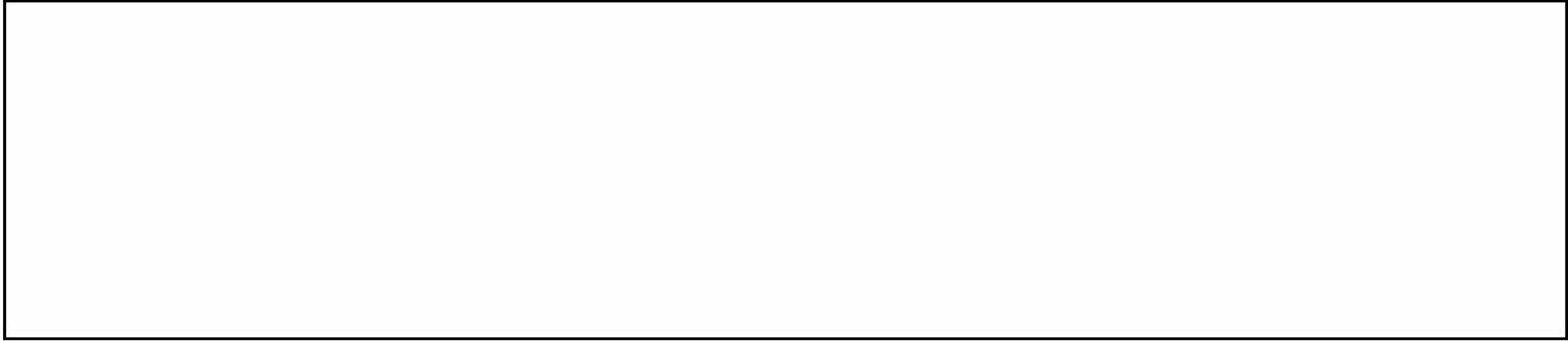


表 4-11　2000—2014 年北京市年用电量指标统计

年份	消费品零售总额/亿元	GDP/亿元	总人口/万人	年用电量/万千瓦小时
2000	1 658.7	3 161.7	1 363.6	3 844 266
2001	1 831.4	3 708.0	1 385.1	3 999 415
2002	2 005.2	4 315.0	1 423.2	4 399 637
2003	2 296.9	5 007.2	1 456.4	4 676 056
2004	2 626.6	6 033.2	1 492.7	5 131 804
2005	2 911.7	6 969.5	1 538.0	5 705 364
2006	3 295.3	8 117.8	1 601.0	6 115 719
2007	3 835.2	9 846.8	1 676.0	6 670 089
2008	4 645.5	11 115.0	1 771.0	6 897 189
2009	5 387.5	12 153.0	1 860.0	7 391 465
2010	6 340.3	14 113.6	1 961.9	8 099 029
2011	7 222.2	16 251.9	2 018.6	8 217 055
2012	8 123.5	17 879.4	2 069.3	8 742 835
2013	8 872.1	19 800.8	2 114.8	9 131 113
2014	9 638.0	21 330.8	2 151.6	9 370 485

资料来源：北京市统计局 2014 统计年鉴，http://www.bjstats.gov.cn/.

7-B1-2　请根据表 4-11 中所列数据，以 GDP 为自变量，年用电量为因变量，利用回归函数，计算截距和斜率，并写出一元线性回归方程。



7-B1-3　请根据表 4-11 中所列数据，以 GDP 为自变量，年用电量为因变量，利用回归工具进行一元线性回归分析。

7-B2-1　请分析“任务 7-B1-3”的输出结果，并预测 GDP 为 22 850.51 亿元时的年用电量，形成文字解释。

7-B2-2　请根据表 4-11 中所列数据，以总人口为自变量，年用电量为因变量，进行回归分析，并预测当总人口为 2 161.7 万时的年用电量，形成文字解释。

7-C1-1　请利用 Excel 求出表 4-12 中所列年销售量依赖地区人口和人均收入的回归方程，进行显著性检验，并形成文字解释。

表 4-12　某电子产品年销售量、地区人口和人均收入数据

编号	年销售量/万台	地区人口/万人	人均收入/万元
1	1.62	27.4	2.450
2	1.20	18.0	3.254
3	2.23	37.5	3.802
4	1.31	20.5	2.838
5	0.67	8.6	2.347
6	1.69	26.5	3.782
7	0.81	9.8	3.008
8	1.92	33.0	2.450
9	1.16	19.5	2.137
10	0.55	5.3	2.560

续表

编号	年销售量/万台	地区人口/万人	人均收入/万元
11	2.52	43.0	4.020
12	2.32	37.2	4.427
13	1.44	23.6	2.660
14	1.03	15.7	2.088
15	2.12	37.0	2.605

7-C1-2　表 4-13 所示是某地区消费者对当地某品牌电子手表的需求量、价格与平均收入的数据。请对需求量随价格与平均收入的变动进行回归分析，并对估计模型进行检验，并形成文字解释。

表 4-13　某地区消费者对当地某品牌电子手表的需求量、价格与平均收入的数据

年份	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
商品价格/(元/件)	89	78	70	60	69	52	45	56	32	45
人均月收入/元	560	530	600	680	750	830	880	830	980	1 100
需求量/件	5 800	4 890	6 200	6 800	7 100	8 900	9 000	8 100	9 990	9 800

7-C2　请根据“任务 7-C1-1”确定的回归方程，预测当地总人口 38 万人、人均收入 4.5 万元时的年销售量。

学习任务页-zh1　财政收入预测问题

学生信息__________　　任务开始日期__________　　任务分值____25 分____

财政收入与国民收入、工业总产值、农业总产值、总人口、就业人口、固定资产投资、财政收入等因素有关。表 4-14 列出了 1978—2014 年的原始数据，试完成以下题目。

(1) 财政收入由国民收入、工业总产值、农业总产值、总人口和固定资产投资等因素决定，试构造财政收入预测模型。

(2) 通过查找资料的方法或预测的方法，补全表 4-14 中所缺的数据，即 2010—2014 年的工业总产值数据。

(3) 2015 年的相关财政数据还未公布，请预测 2015—2016 年的财政收入。

表 4-14　1978—2014 年国民收入、工业总产值、农业总产值、总人口、固定资产投资、财政收入原始数据

年份	国民收入/亿元	工业总产值/亿元	农业总产值/亿元	总人口/万人	固定资产投资/亿元	财政收入/亿元
1978	3 645	1 745	1 026	96 259	550	1 132
1979	4 063	1 914	1 270	97 542	564	1 146
1980	4 546	2 192	1 372	98 705	568	1 160
1981	4 890	2 256	1 560	100 072	961	1 176
1982	5 331	2 383	1 777	101 654	1 200	1 212
1983	5 986	2 646	1 978	103 008	1 369	1 367
1984	7 244	3 106	2 316	104 357	1 833	1 643
1985	9 041	3 867	2 564	105 851	2 543	2 005
1986	10 274	4 493	2 789	107 507	3 121	2 122
1987	12 051	5 252	3 233	109 300	3 792	2 199
1988	15 037	6 587	3 865	111 026	4 754	2 357
1989	17 001	7 278	4 266	112 704	4 410	2 665
1990	18 718	7 717	5 062	114 333	4 517	2 937
1991	21 826	9 102	5 342	115 823	5 595	3 149
1992	26 937	11 700	5 867	117 171	8 080	3 483
1993	35 260	16 454	6 964	118 517	13 072	4 349
1994	48 109	22 445	9 573	119 850	17 042	5 218

续表

年份	国民收入/亿元	工业总产值/亿元	农业总产值/亿元	总人口/万人	固定资产投资/亿元	财政收入/亿元
1995	59 811	28 680	12 136	121 121	20 019	6 242
1996	70 143	33 835	14 015	122 389	22 914	7 408
1997	78 061	37 543	14 442	123 626	24 941	8 651
1998	83 024	39 004	14 818	124 761	28 406	9 876
1999	88 479	41 034	14 770	125 786	29 855	11 444
2000	98 001	45 556	14 945	126 743	32 918	13 395
2001	108 068	49 512	15 781	127 627	37 214	16 386
2002	119 096	53 897	16 537	128 453	43 500	18 904
2003	135 174	62 436	17 382	129 227	55 567	21 715
2004	159 587	73 904	21 413	129 988	70 477	26 396
2005	184 089	87 365	22 420	130 756	88 774	31 649
2006	213 132	103 162	24 040	131 448	109 998	38 760
2007	268 631	124 799	28 627	132 129	137 324	51 322
2008	318 737	146 183	34 000	132 802	172 828	61 330
2009	345 046	156 958	35 226	133 450	224 599	68 477
2010	407 138	—	36 941	134 091	251 684	83 102
2011	479 576	—	41 989	134 735	311 485	103 874
2012	532 872	—	46 940	135 404	374 695	117 254
2013	583 198	—	51 497	136 072	446 294	129 209
2014	634 043	—	54 772	136 782	512 021	140 370

资料来源：国家统计局，http://www.stats.gov.cn/.

学习任务页-zh2　销售预测分析

学生信息＿＿＿＿＿　　任务开始日期＿＿＿＿＿　　任务分值＿25 分＿

"阿曼德比萨"是一个制作和外卖意大利比萨的餐饮连锁店，其主要客户群是在校大学生。为了研究各店铺销售额与店铺附近大学生人数之间的关系，随机抽取十个分店的样本，得到的数据如表 4-15 所示。

表 4-15　十个分店的样本数据

店铺编号	区内大学生人数 X/万人	季度销售额 Y/万元
1	0.2	5.8
2	0.6	10.5
3	0.8	8.8
4	0.8	11.8
5	1.2	11.7

续表

店铺编号	区内大学生人数 X/万人	季度销售额 Y/万元
6	1.6	13.7
7	2.0	15.7
8	2.0	16.9
9	2.2	14.9
10	2.6	20.2

对数据进行回归分析，并预测区内大学生人数为 1.8 万人的店铺 11 的季度销售额。

学习信息库 1——回归分析概述

相关分析是用相关系数或相关指数等指标分析变量之间相互依存的密切程度及相关的方向，那么，如何分析变量之间具体的数量变动关系呢？这就是回归分析问题。

1. 回归分析的基本概念

(1) 回归分析

研究具有相关关系的变量之间的数量变动关系，由一个或多个变量来估计或预测另一个变量的值时，所建立的数学模型叫作回归模型；所进行的统计分析，称为回归分析。

回归分析，要以现象之间存在相关关系为前提，然后对自变量和因变量的变动拟合适宜的回归方程，确定其定量关系式，再对拟合的回归方程进行显著性检验，最后利用求得的关系式进行推算和预测。

(2) 回归分析的意义和任务

回归分析的主要意义和目的是进行估算或预测。对具有相关关系的两个或两个以上变量之间数量变化的一般关系进行测定，确立一个相应的数学表达式，以便从已知量推测未知量，为估算预测提供重要的方法。

回归分析的主要任务是根据估算或预测目标，建立适合回归总体的样本回归函数。

(3) 回归的类型

按自变量的个数，分为一元回归和多元回归；按回归线的形态，分为线性回归和非线性回归，如表 4-16 所示。

表 4-16　回归的类型

标签	自变量的个数	回归线的形态
类别	① 一元回归 回归分析中,只有一个自变量,即用一个自变量预测因变量 ② 多元回归 回归分析中,有多个自变量,即用多个自变量预测因变量	① 线性回归 回归线表现为直线形式,即自变量和因变量之间线性相关 ② 非线性回归 回归线表现为曲线形式,即自变量和因变量之间非线性相关

(4) 回归分析和相关分析的联系和区别

① 联系。相关程度越高,回归方程越可靠。相关系数和回归系数方向一致,可以互相推算。两者都是研究变量之间相互关系的基本统计方法,它们是相互补充的关系。回归分析是在相关分析的基础上进行,没有相关分析的回归分析就像“无头有尾”;没有回归分析的相关分析就像“有头无尾”。

② 区别。两者的概念和作用不同,它们从不同的角度说明现象之间的依存关系。相关分析只能说明现象之间相关的方向和密切程度,不能说明一个现象随着另一个现象的变化而发生多大的变化。回归分析通过建立适宜的回归方程,能够测出这种变化的量,它是进行推算和预测的重要依据。相关分析不用确定自变量和因变量,而回归分析需要确定自变量和因变量。

2. 回归分析的基本步骤

回归分析一般按以下五个步骤操作。

(1) 根据预测目标,确定自变量和因变量

明确预测的具体目标,也就确定了因变量。如预测具体目标是下一年度的销售量,那么销售量 Y 就是因变量。通过市场调查和查阅资料,寻找与预测目标相关的影响因素,即自变量,从中选出主要的影响因素。

(2) 进行相关分析

通过绘制变量之间的散点图,初步判断变量之间的相关性并做出相关分析。变量之间存在相关关系是回归分析的基础。只有当变量与因变量确实存在某种关系时,建立的回归方程才有意义。

(3) 确定回归方程

定量分析变量之间的关系,通过最小二乘法估计回归系数,确定回归方程;或者用计算机软件确定回归方程,比如 Excel。

(4) 回归的显著性检验

回归预测模型是否可用于实际预测,取决于对回归预测模型的检验和对预测误差的计算。回归方程只有通过各种检验,且预测误差较小,才能将回归方程作为预测模型进行预测。

(5) 回归预测,确定预测值

利用回归模型计算,确定预测值。

3. 一元线性回归

一元线性回归是只有一个自变量,且自变量和因变量之间是线性相关关系的回归分析,

是最简单的回归模型。一元线性回归模型虽然简单,但它是更复杂类型回归分析的基础,其建立模型的过程、基本思想以及在实际问题中的应用原理都是相通的。

(1) 一元线性回归模型

① 总体回归模型。当两个变量之间存在显著的相关关系时,可对其进行一元线性回归分析,所建立的模型称为一元线性回归模型,简称简单线性回归模型。在该模型中,假设因变量 Y 的变化主要受自变量 X 的影响,它们之间存在大致的线性函数关系,表示为

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{4-1}$$

式(4-1)称为总体回归模型。其中,β_0 和 β_1 是待定参数,称为回归系数;Y_i 和 X_i 分别是因变量 Y 和自变量 X 的第 i 个观测值;u_i 是随机误差项,反映建立模型时未考虑进去的其他各种因素对 Y 的影响。u_i 是无法直接观测的,理论上假设其数学期望值为零,方差为常数。

② 样本回归模型。在实际研究中,很多现象的总体一般很多,通常情况下难以获得,或者费时费力。因此,总体回归模型难以得到,需要利用一定数量的样本数据对其进行估计。

根据样本数据拟合的直线,称为样本回归直线。在两个变量相关的散点图中,引出一条最优的直线,这条直线就是样本回归直线。一元线性回归模型的样本回归线表示为

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \tag{4-2}$$

式(4-2)称为样本回归函数。其中,$\hat{\beta}_0$ 和$\hat{\beta}_1$ 是待定参数,$\hat{\beta}_0$ 称为截距,$\hat{\beta}_1$ 称为斜率,表示自变量变动一个单位时,因变量平均变动的量,它们是总体回归系数 β_0 和 β_1 的估计值。$\hat{Y}_i$ 是样本回归线上与 X_i 对应的因变量 Y 的估计值,但不一定等于实际观测值 Y_i,它们之间存在一定差距。

令 $e_i = Y_i - \hat{Y}_i$,则有

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \quad (i = 1,2,\cdots,n) \tag{4-3}$$

式(4-3)称为样本回归模型。其中,e_i 称为残差;n 是样本容量,即样本个数。

(2) 一元线性回归模型回归系数的估计

① 回归系数的点估计。回归分析就是通过样本得到回归总体的样本回顾模型,以此来估算或预测。只要确定了回归系数$\hat{\beta}_0$ 和$\hat{\beta}_1$,即可确定样本回归模型。在实际问题中,确定样本回归模型时,要掌握估计值尽量接近实际观测值或真实值的原则,即残差的总体影响越小越好。那么,如何衡量残差的总体影响呢? 由于残差的值有正有负,若直接求和,可能导致相互抵消而忽略这部分残差的影响,因此采用残差平方和来估计所有残差的总体影响,可以根据残差平方和最小的原则来估计样本回归系数,这就是所谓的最小二乘法。

令

$$Q(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} e_i^2 \tag{4-4}$$

由式(4-3)可知

$$Q(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \tag{4-5}$$

根据微积分中求极值与最值的原理，要得到二元函数 $Q(\hat{\beta}_0,\hat{\beta}_1)$ 的最小值，$Q(\hat{\beta}_0,\hat{\beta}_1)$ 对 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的偏导数必须为零，即

$$\frac{\partial Q}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \tag{4-6}$$

$$\frac{\partial Q}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n}X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \tag{4-7}$$

整理式(4-6)和式(4-7)，有

$$n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n}X_i = \sum_{i=1}^{n}Y_i \tag{4-8}$$

$$\hat{\beta}_0\sum_{i=1}^{n}X_i + \hat{\beta}_1\sum_{i=1}^{n}X_i^2 = \sum_{i=1}^{n}X_iY_i \tag{4-9}$$

联立式(4-8)和式(4-9)，求得

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n}X_iY_i - \sum_{i=1}^{n}X_i\sum_{i=1}^{n}Y_i}{n\sum_{i=1}^{n}X_i^2 - \left(\sum_{i=1}^{n}X_i\right)^2} \tag{4-10}$$

$$\hat{\beta}_0 = \sum_{i=1}^{n}\frac{Y_i}{n} - \hat{\beta}_1\sum_{i=1}^{n}\frac{X_i}{n} = \bar{Y} - \hat{\beta}_1\bar{X} \tag{4-11}$$

式(4-10)和式(4-11)就是总体回归系数 β_0 和 β_1 的点估计公式。

例 4-1 表 4-17 所示是生产费用(万元)和月产量(吨)的相关数据。根据数据，求出回归方程，并说明回归系数的含义。

解 根据表 4-17 中所列数据和式(4-10)、式(4-11)，有

$$\sum x = 36.4,\quad \sum y = 880,\quad n = 8,\quad \sum x^2 = 207.54$$

$$\sum y^2 = 104\ 214,\quad \sum xy = 4\ 544.6$$

$$\hat{\beta}_1 = \frac{n\sum xy - \sum x\sum y}{n\sum x^2 - \left(\sum x\right)^2} = 12.9$$

$$\hat{\beta}_0 = \frac{\sum y}{n} - b_1\frac{\sum x}{n} = 51.3$$

由式(4-2)，可得回归方程为

$$\hat{y} = 51.3 + 12.9x$$

回归系数 $\hat{\beta}_1$ 的含义：月产量每增加 1 000 吨，生产费用平均增加 12.9 万元。

表 4-17 生产费用和月产量的相关数据(求回归方程)

序号	月产量 x	生产费用 y	x^2	y^2	xy
1	1.2	62	1.44	3 844	74.4
2	2.0	86	4.00	7 396	172.0
3	3.1	80	9.61	6 400	248.0

续表

序号	月产量 x	生产费用 y	x^2	y^2	xy
4	3.8	110	14.44	12 100	418.0
5	5.0	115	25.00	13 225	575.0
6	6.1	132	37.21	17 424	805.2
7	7.2	135	51.84	18 225	972.0
8	8.0	160	64.00	25 600	1 280.0
$\sum$	36.4	880	207.54	104 214	4 544.6

例 4-2 某地区人均年收入与耐用消费品销售额数据如表 4-18 所示。

表 4-18 某地区人均年收入与耐用消费品销售额数据

时间	人均年收入/元	耐用消费品销售额/万元
1990	340	82
1991	380	90
1992	450	100
1993	470	114
1994	560	140
1995	620	144
合计	2 820	670

请分析两个变量相关的密切程度。若为显著相关以上，对两个变量进行回归分析。

解 根据相关系数公式(3-2)及式(4-10)、式(4-11)，有

$$r = 0.98,\quad \hat{\beta}_0 = -1.13,\quad \hat{\beta}_1 = 0.24$$

则由式(4-2)，可知回归方程为

$$\hat{y} = -1.13 + 0.24x$$

回归系数$\hat{\beta}_1$ 的含义：人均年收入每增加 1 元，耐用消费品销售额平均增加 0.24 万元。

② 回归系数的区间估计。利用回归系数的点估计式(4-10)和式(4-11)及其方差，得到回归系数变动的范围，即回归系数的区间估计公式。

利用 Excel 做回归分析，可轻松得到回归系数的区间估计。因此，关于区间估计的原理及其公式的推导过程，这里不详细介绍，可参考其他统计学书籍。

(3) 一元线性回归模型的检验

得到回归模型后，不能直接用来估算或预测，必须进行检验。若检验不通过，则需回到模型建立的前几个环节检查根源并调整模型。

① 回归模型检验的种类。回归模型的检验分以下三大类。

第一类：现实意义检验。

现实意义检验主要考察回归系数的符号及取值范围是否符合所讨论问题的实质性理论和实践经验。

第二类：拟合优度评价、显著性检验。

这两个检验又称统计学检验，它们通过统计学的抽样理论来检验样本回归模型的可靠性，是对所有现象的回归分析必须通过的检验。

第三类：序列相关检验、异方差检验、多重共线性检验。

这三个检验对于社会经济现象的定量分析具有特别重要的意义。关于这类检验，可参考其他相关书籍。

② 拟合优度的评价。拟合优度评价指评价回归线拟合的好坏程度。

首先，可从添加回归直线的散点图直观评价，观察样本观测点在样本回归线周围的聚集程度。

其次，根据样本决策系数(简称决策系数)来评价，一般用 r^2 或 R^2 表示。在一元线性回归中，决定系数就是相关系数的平方。它是回归模型拟合优度评价的综合尺度，其取值范围为[0,1]。决定系数越大，模型拟合优度越好；决定系数越小，模型拟合优度越差。

③ 回归的显著性检验。回归的显著性检验包括两个方面，一是对回归系数的显著性检验；二是对整个回归方程的显著性检验。

回归系数的检验可采用 t 值检验或 P 值检验，回归方程的检验可采用 F 值检验或 F 显著性水平检验。

t 和 F 均为统计量，t 值检验和 F 值检验都需要查询它们的临界值并进行比较，用起来较麻烦；P 值检验和 F 显著性水平检验不需要查询临界值，直接与显著性水平 α 相比较即可。若 P 值小于 α，对应回归系数的检验通过，反之不然。若 F 显著性水平小于 α，则回归方程检验通过，反之不然。用 Excel 做回归分析时，这些值都可以一次性获得，所以关于它们的公式及原理，这里不详细介绍，可参考其他书籍。

(4) 一元线性回归模型的预测

建立回归模型的主要目的是为了估算或预测。当所建立的回归方程通过检验后，可用其来进行预测。

简单回归预测就是把给定的解释变量(自变量)的值代入通过检验的回归函数式(4-2)，并计算对应的函数值，称为点预测，其公式为

$$\hat{Y}_m = \hat{\beta}_0 + \hat{\beta}_1 X_m \tag{4-12}$$

其中，X_m 是给定自变量的值；$\hat{Y}_m$ 是对应于 X_m 的因变量的预测值。

一元线性回归预测有以下两种情况。

① 当给定的 X_m 是样本区间之内的值时，该预测叫作插值。

② 当给定的 X_m 是样本区间之外的值时，该预测叫作外推。

4. 多元线性回归

一元线性回归研究一个因变量与一个自变量之间的关系。但在现实中，需要考虑多个因素对一个现象的影响。

在线性相关条件下，研究两个或两个以上自变量对一个因变量的数量变化关系，称为多元线性回归分析，所建立的模型称为多元线性回归模型。

多元线性回归分析是一元线性回归分析的推广，基本原理与一元回归模型基本类似，只是计算比较麻烦，本书不详细介绍。

多元线性回归模型的总体回归函数的一般形式为

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_i X_{ji} + \cdots + \beta_n X_{ni} + u_i \tag{4-13}$$

式(4-13)是因变量 Y 与 n 个自变量 $X_1, X_2, \cdots, X_n$ 之间的多元线性回归模型，X_{ji} 表示第 j 个自变量 $X_j(j=1,2,\cdots,n)$ 的第 i 个观测值，u_i 是随机误差项，$\beta_1, \beta_2, \cdots, \beta_n$ 是总体回归系数。

多元线性回归模型的样本回归函数为

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_n X_{ni} \tag{4-14}$$

多元线性回归模型的样本回归模型为

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_j X_{ji} + \cdots + \hat{\beta}_n X_{ni} + e_i \tag{4-15}$$

式(4-15)称为多元线性回归样本回归模型。其中，$\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_n$ 是待定的样本回归系数，它们是总体回归系数 $\beta_1, \beta_2, \cdots, \beta_n$ 的估计值；$\hat{\beta}_j$ 表示在其他自变量保持不变的情况下，自变量 X_j 变动一个单位所引起的因变量 Y 平均变动的量；e_i 是残差。

注意：在多元回归模型中，假设包含的自变量之间不能具有较强的线性关系。

多元线性回归模型的估计、检验和预测类似于一元线性回归模型的情形，本书不详细介绍。

5. 非线性回归分析

当自变量与因变量之间呈现为非线性关系时，需要做非线性回归分析，以便更准确地反映与所研究变量之间的关系。与线性回归不同，非线性回归函数有多种不同的形式，需要根据所讨论的问题恰当选择。一般情况下，把非线性回归问题转化为线性回归问题再处理，但并不是所有非线性函数都可以转化为线性函数。下面介绍几种把非线性函数转化为线性函数的常用方法。

(1) 倒数变换

倒数变换是把模型中自变量的倒数换元为新的变量，使原非线性函数转化成新的变量的线性函数。例如，对于双曲线函数

$$Y = a + \frac{b}{X} \tag{4-16}$$

令 $X' = \frac{1}{X}$，代入式(4-16)，得到线性函数 $Y = a + bX'$。

(2) 单对数变换

单对数变换主要用于对数函数的线性变换，把原函数中自变量的对数部分换元为新的变量，从而把对数函数转化为线性函数。例如，对于对数函数

$$Y = a + b\ln X \tag{4-17}$$

令 $X' = \ln X$，代入式(4-17)，得到线性函数 $Y = a + bX'$。

(3) 多对数变换

多对数变换主要用于指数函数和幂函数的线性转换，先对原函数两边同时取对数，再把函数中的对数部分换元为新的变量或参数，从而把原函数转化为线性函数。例如，对于指数函数

$$Y = ab^X \tag{4-18}$$

式(4-18)两边求对数，有

$$\ln Y = \ln a + X\ln b \tag{4-19}$$

令 $Y'=\ln Y$，$a'=\ln a$，$b'=\ln b$，代入式(4-19)，得到线性函数 $Y'=a'+b'X$。

对于幂函数

$$Y = aX^b \tag{4-20}$$

式(4-20)两边求对数，有

$$\ln Y = \ln a + b\ln X \tag{4-21}$$

令 $Y'=\ln Y$，$a'=\ln a$，$X'=\ln X$，代入式(4-21)，得到线性函数 $Y'=a'+bX'$。

至于如何选择合适的非线性回归模型形式，有以下三种方法。

① 根据理论和经验或观察散点图，初步选择非线性回归模型形式。

② 模型要有较好的拟合度。若拟合度不理想，再调整模型。

③ 函数的数学形式尽可能简单。若几种函数都符合要求，应选择数学形式较简单的一种。

学习信息库 2——利用 Excel 做回归分析

利用 Excel 软件，可以较轻松地完成一套完整的回归分析工作。下面结合实例介绍三种与回归分析有关的 Excel 功能。

1. 利用图表分析

图表分析方法适合利用 Excel 做简单一元线性回归，操作步骤如下所述。

首先，制作相应数据的散点图。

其次，右击散点图，添加趋势线，如例 4-3 所示。

例 4-3 2000—2014 年北京市年用电量统计数据截图如图 4-1 所示。请以年用电量为因变量 y，以总人口为自变量 x，确定一元线性回归函数。

	A	B	C	D	E
1	年份	消费品零售总额/亿元	GDP/亿元	总人口/万人	年用电量/万千瓦小时
2	2000	1658.7	3161.7	1363.6	3844266
3	2001	1831.4	3708.0	1385.1	3999415
4	2002	2005.2	4315.0	1423.2	4399637
5	2003	2296.9	5007.2	1456.4	4676056
6	2004	2626.6	6033.2	1492.7	5131804
7	2005	2911.7	6969.5	1538.0	5705364
8	2006	3295.3	8117.8	1601.0	6115719
9	2007	3835.2	9846.8	1676.0	6670089
10	2008	4645.5	11115.0	1771.0	6897189
11	2009	5387.5	12153.0	1860.0	7391465
12	2010	6340.3	14113.6	1961.9	8099029
13	2011	7222.2	16251.9	2018.6	8217055
14	2012	8123.5	17879.4	2069.3	8742835
15	2013	8872.1	19800.81	2114.8	9131113
16	2014	9638.0	21330.83	2151.6	9370485

图 4-1　2000—2014 年北京市年用电量统计数据截图

解　首先，在 Excel 表中选中所需数据，即总人口及年用电量两列数据 \$D\$1:\$E\$16，插入散点图，如图 4-2 所示。

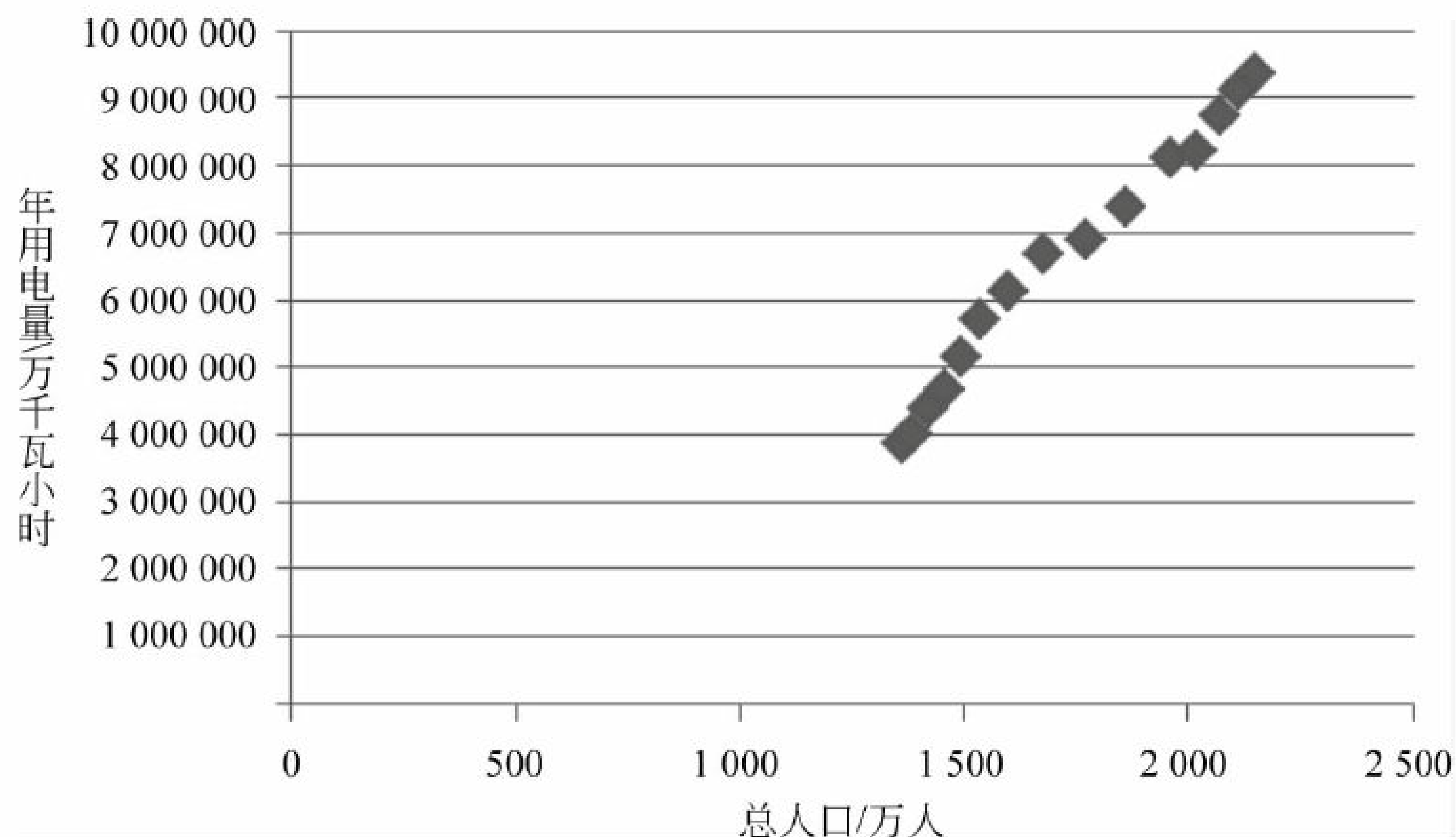


图 4-2　总人口与年用电量的散点图

其次，右击散点图中的数据标记，选择"添加趋势线"，弹出"设置趋势线格式"对话框，如图 4-3 所示。

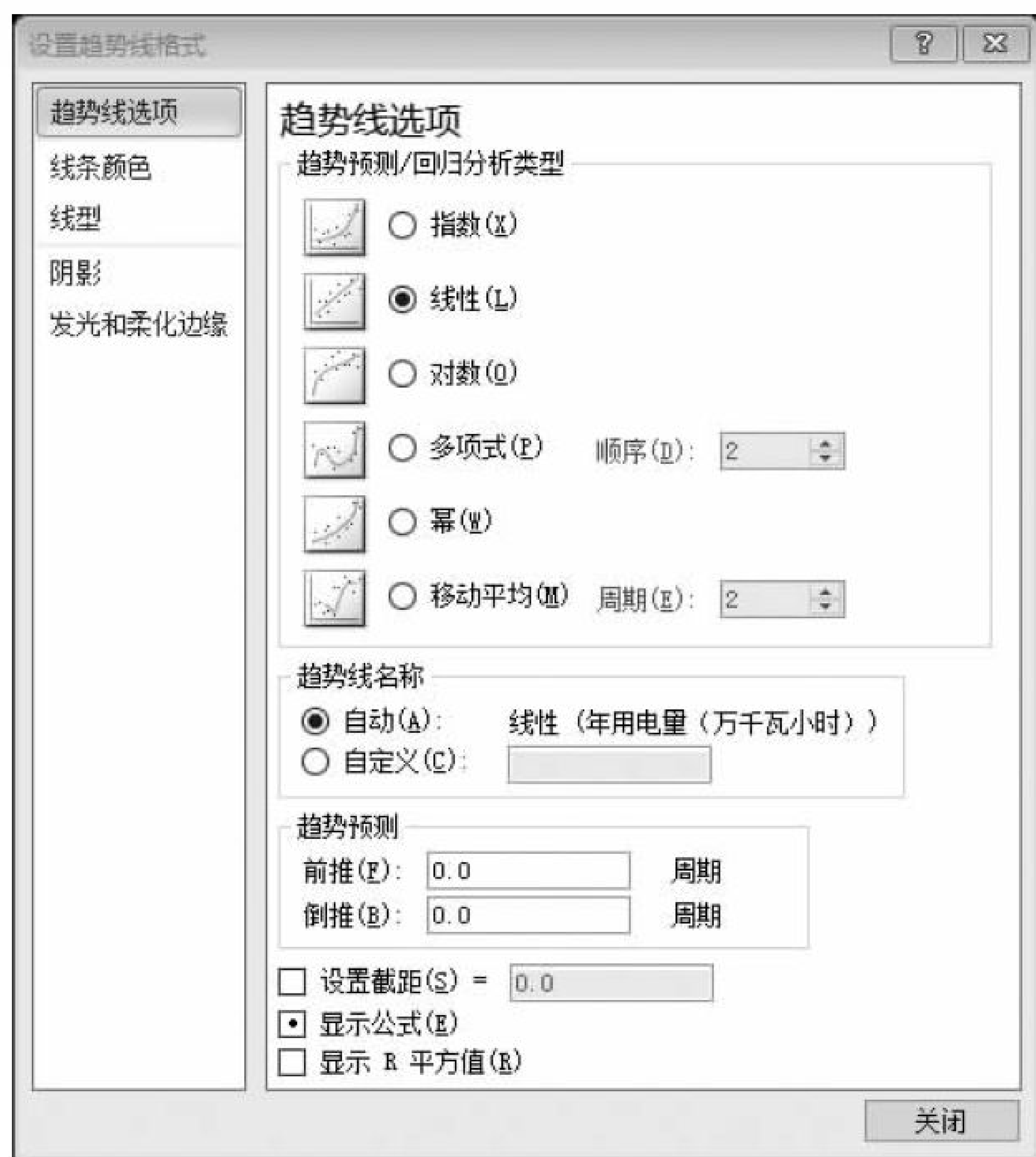


图 4-3　"设置趋势线格式"对话框

最后，在"设置趋势线格式"对话框中进行相应的设置。"趋势线选项"选择"线性"，勾选"显示公式"和"显示 R 平方值"，其他选择默认项。最后，单击"关闭"按钮，得到带有趋势线（回归直线）和回归函数的散点图，如图 4-4 所示。

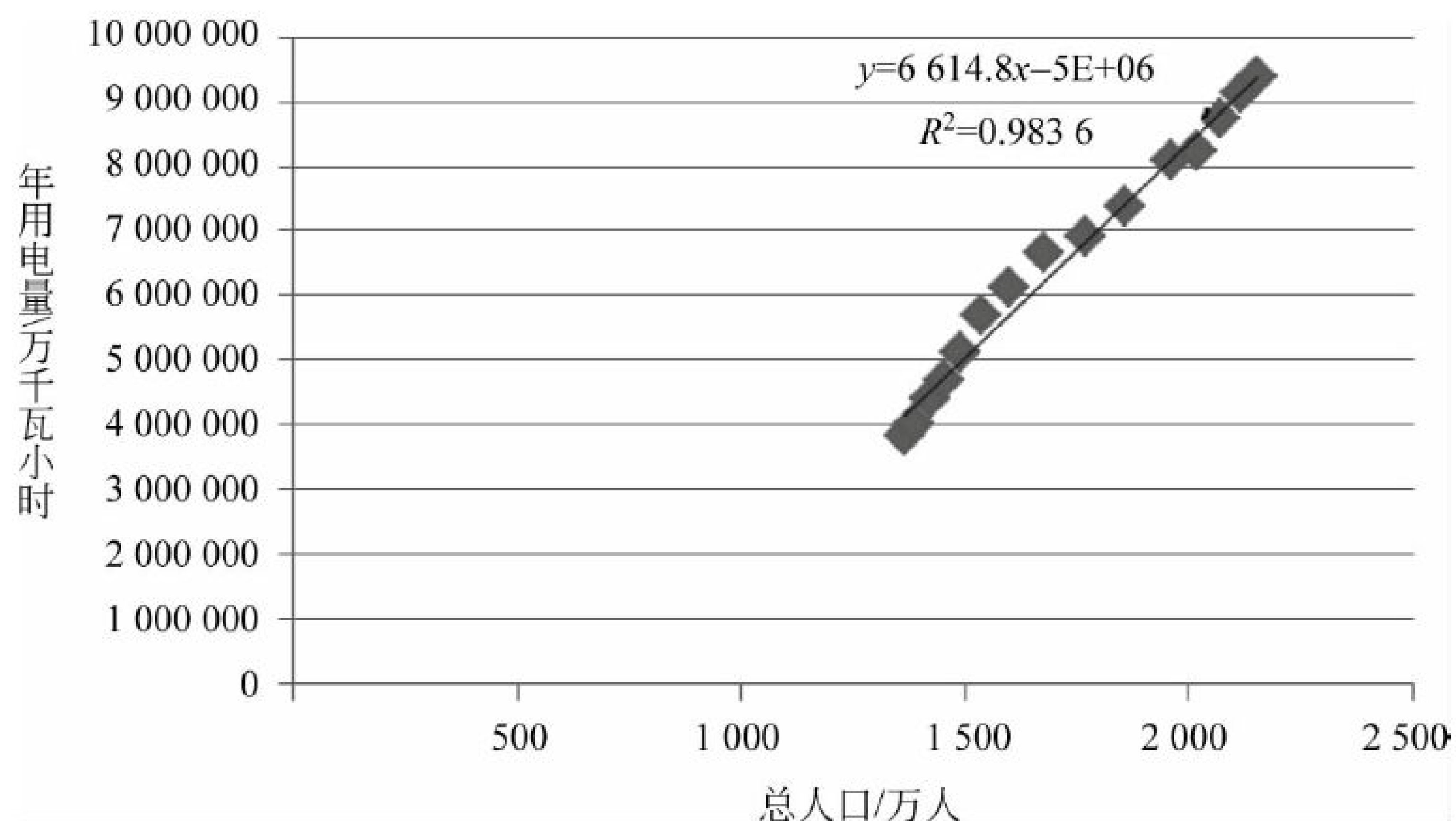


图 4-4　带有趋势线和回归函数的散点图

在图 4-4 中,"−5E+06"是科学记数法,等于"−5 乘以 10 的 6 次幂",即年用电量与总人口的回归函数为

$$y = -5\ 000\ 000 + 6\ 614.8x \tag{4-22}$$

决定系数 $R^2 = 0.983\ 6$,说明拟合优度很好。

2. 利用回归函数分析

Excel 提供了完整的统计分析函数。这里简单介绍一元线性回归的截距函数和斜率函数,更多有关回归分析函数的应用可参考其他相关书籍。

(1) 截距函数:INTERCEPT。

语法:INTERCEPT(Known_y's,Known_x's)

其中,Known_y's 为因变量的数据区域;Known_x's 为自变量的数据区域。

(2) 斜率函数:SLOPE。

语法:SLOPE(Known_y's,Known_x's)

其中,Known_y's 为因变量的数据区域;Known_x's 为自变量的数据区域。

例 4-4　请利用回归函数计算例 4-3 中的截距与斜率。

解　第一步:在 Excel 工作表中依次单击"公式"→"插入函数"→"统计",找到 INTERCEPT 或 SLOPE 并确定,打开"函数参数"对话框,如图 4-5 和图 4-6 所示。



图 4-5　INTERCEPT 对话框

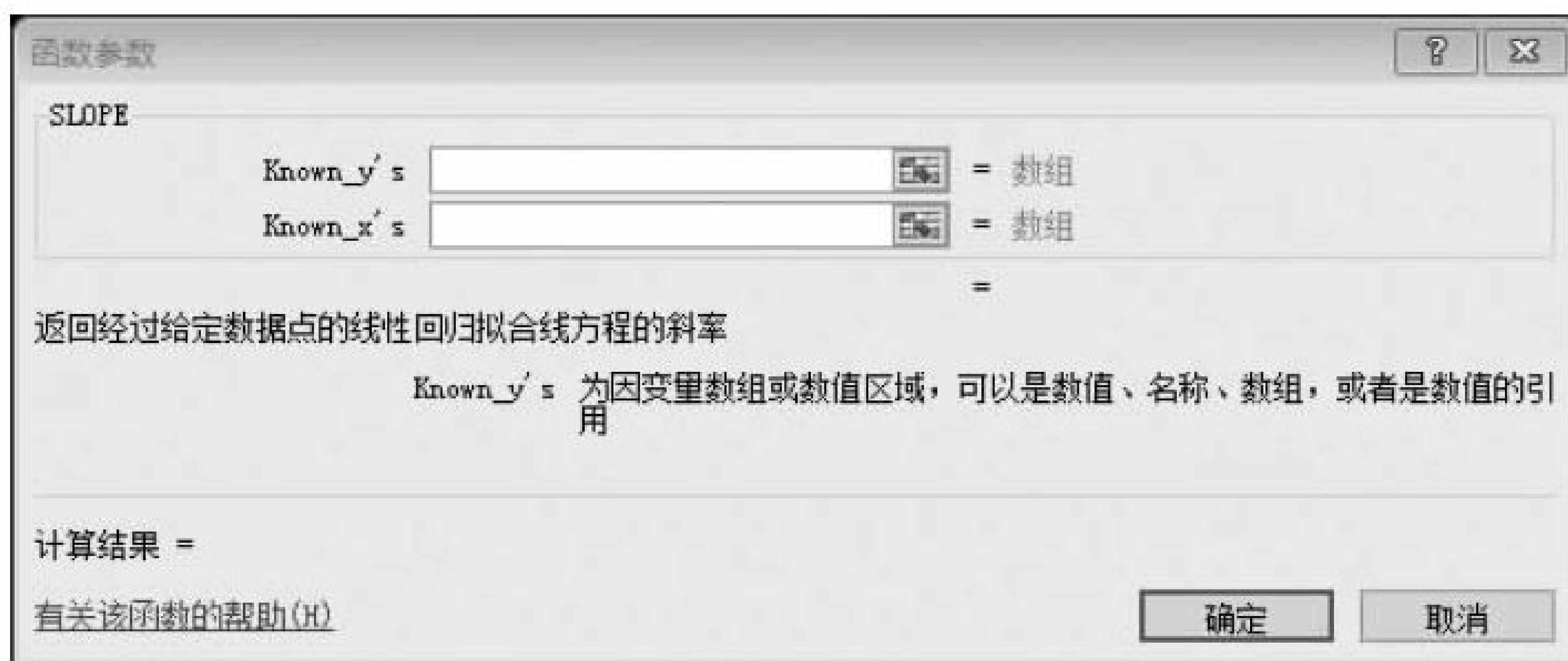


图 4-6　SLOPE 对话框

第二步：在相应区域输入对应的数值区域，然后单击“确定”按钮。因变量 y 与自变量 x 按照图 4-7 和图 4-8 所示输入。



图 4-7　输入数值区域的 INTERCEPT 对话框



图 4-8　输入数值区域的 SLOPE 对话框

此时计算出截距 $\beta_0 = -4\ 854\ 648.297$ 和斜率 $\beta_1 = -6\ 614.76$。因此，一元线性回归函数为

$$y = -4\ 854\ 648.297 + 6\ 614.76x \tag{4-23}$$

3. 利用回归分析工具分析

上面两种方法用起来均比较方便，但只能完成简单回归分析工作，后续的回归检验所需的 t 统计量及 P 值，F 统计量及 F 显著性水平等值都未得到。Excel 分析工具库提供了完整的回归工具，能够一次性算出包括各项回归检验指标的完整信息。默认状态下，Excel 不显示回归工具，首次使用时需要安装。

例 4-5 请根据表 4-11 所列数据完成以下任务。

(1) 确定以年用电量为因变量,以消费品零售总额、GDP 与总人口为自变量的三元线性回归函数,并进行线性回归检验。

(2) 假设消费品零售总额为 9 850 亿元,GDP 为 22 850.51 亿元,总人口为 2 161.7 万人,预测年用电量。

解 第一步:在 Excel 工作表中依次单击"数据"→"数据分析"→"回归"→"确定",如图 4-9 所示,调出空白的"回归"对话框,如图 4-10 所示。



图 4-9 "数据分析"窗口

第二步:在空白的"回归"对话框中输入相应的数据,进行相应的设置,如图 4-11 所示。单击"确定"按钮,得到回归分析结果,如图 4-12 所示。



图 4-10 空白的"回归"对话框



图 4-11 设置好的"回归"对话框

SUMMARY OUTPUT								
回归统计								
Multiple R	0.9979019							
R Square	0.9958083							
Adjusted R	0.9946651							
标准误差	138476.61							
观测值	15							
方差分析								
	df	SS	MS	F	Significance F			
回归分析	3	5.01E+13	1.67E+13	871.0753	2.3602E-13			
残差	11	2.11E+11	1.92E+10					
总计	14	5.03E+13						
	Coefficient	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
Intercept	-3114989	1235062	-2.52213	0.028368	-5833341.91	-396637.1	-5833342	-396637.1
消费品零售总	-941.5707	168.8312	-5.577	0.000166	-1313.165597	-569.9758	-1313.166	-569.9758
GDP	490.56051	92.58712	5.298367	0.000253	286.7776372	694.34339	286.77764	694.34339
总人口	5149.3898	1002.94	5.134294	0.000326	2941.933249	7356.8464	2941.9332	7356.8464

图 4-12 "回归分析"输出结果

注意：

(1) 使用“回归”工具时，所有变量必须按列存放。

(2)“回归”对话框可以按照不同的目的及要求灵活设置。

① Y 值输入区域：因变量数据区域。该区域只能是单列数据。

② X 值输入区域：自变量数据区域。该区域只能是单列数据。

③ 标志：若输入的数值区域包括变量名称，则勾选此项。

④ 常数为零：若回归函数中不想包括常数项，则勾选此项。

⑤ 置信度：若需要置信度信息，则勾选此项。

其他选项都类似。若在输出表中需要这些信息，就选择；不需要，就不用选择。

回归分析输出结果解读：Excel 回归分析工具的输出信息包括 3 个部分，分别为回归统计部分、方差分析部分和回归参数部分。

(1) 回归统计部分

回归统计表包括以下几部分内容。

① Multiple R：复相关系数，用来衡量变量之间的相关程度。本例中，R 为 0.99，说明变量之间高度相关，满足回归分析的前提条件。

② R Square(R 平方)：决定系数，用来衡量回归拟合的优度。决定系数越大，说明回归函数的拟合优度越好。本例中，决定系数等于 0.99，所以拟合优度很好。

③ Adjusted R Square：修正的决定系数，仅用于说明多元回归的拟合优度，其值越大，拟合优度越好。

④ 准误差：用来衡量拟合优度，也用于计算与回归相关的其他统计量。此值越小，拟合优度越好。

⑤ 观测值：样本观测值个数。

(2) 方差分析部分

方差分析的主要作用是通过 F 检验来检验回归模型的整体回归效果，主要参数如下所述。

df：自由度。

SS：离差的平方和。

MS：均方差。

F：F 统计量。本题中等于 871.075 3。

Significance F：F 显著性水平。本题中等于 2.360 2E−13。

(3) 回归参数部分

回归参数部分主要用于回归系数的估计和回归参数的推断，主要参数如下所述。

Intercept：回归截距。

Coefficients：回归系数。

t Stat：t 统计量。

P-value：P 值。

设因变量年用电量为 Y，消费品零售总额为 X_1，自变量 GDP 为 X_2，总人口为 X_3，从输出结果(见图 4-13)可知

$$\beta_0 = -3\ 114\ 989,\quad \beta_1 = -941.57,\quad \beta_2 = 490.56,\quad \beta_3 = 5\ 149.39$$

则回归函数为

$$Y = -3\ 114\ 989 - 941.57X_1 + 490.56X_2 + 5\ 149.39X_3 \qquad (4\text{-}24)$$

接下来从以下三方面做显著性检验。

首先，按照经验初步检验各回归系数的符号是否符合理论意义。比如，本题中总人口的回归系数为 490.56，由回归系数的含义可知，当总人口增加 1 万人，年用电量增加 490.56 千瓦时。类似地，可初步判断各系数的符号符合理论意义。

其次，回归函数的整体回归效果检验。采用 F 检验。根据 4.3.3 小节的介绍，用 F 显著性水平检验更方便，且结论一致。本题中，F 显著性水平为 2.360 2E－13，若显著性水平 $\alpha=0.05$，则 2.360 2E－13 远远小于 0.05，说明回归函数的整体回归效果通过显著性检验。

最后，各回归系数的显著性检验。可采用 t 检验和 P 值检验。两种方法的检验结果是一致的，但 P 值检验用起来更方便。本题中，三个回归系数的 P 值分别为

$$P_1 = 0.000\ 166, \quad P_2 = 0.000\ 253, \quad P_2 = 0.000\ 326$$

都小于显著性水平 0.05，能通过显著性检验。

综合上述分析，认为建立的回归模型通过检验，可用其来进行预测。到此，例题的第一问解决完毕，接下来解决第二问。

根据题意，当 $X_1=9\ 850$，$X_2=22\ 850.51$，$X_3=2\ 161.7$ 时，预测年用电量 Y。这属于点预测问题，直接代入回归模型(4-23)中计算，得到此时的年用电量预测值为9 951 529.049 千瓦时。

这个计算也可用 Excel 的公式完成。把所需数据输入 Excel 工作表，再用公式计算，如图 4-13 所示。

E2 fx =D2+C2*B2+C3*B3+C4*B4

	A	B	C	D	E
1	自变量	给定值	回归系数	常数项	因变量的预测值
2	消费品零售总额（亿元）	9850	-941.57	-3114989	9951529.049
3	GDP （亿元）	22850.51	490.56		
4	总人口 （万人）	2161.7	5149.39		

图 4-13 计算预测值截图

学习信息库 3——优秀论文

牙膏销售量的回归预测模型

摘要 文章应用回归分析理论及机理分析方法讨论了牙膏销售量与销售价格、广告投入费用之间的关系。

牙膏是生活必需品，在合理定价范围之内无论其价格有多高，总会有需求。因此，人们关注更多的是同类商品之间的价格差，而不是价格本身。以回归分析理论为依据，以 Excel 软件为工具，得到了牙膏销售量与同类商品价格差和广告投入费用的关系为高度正线性相关关系，相关系数 $r=0.94$，得到了回归预测模型如下：

$$\hat{Q} = 4.41 + 1.59p_1 + 0.56p_2$$

当价格差为 0.38 元，投入 6.6 百万元的广告费用时，该公司牙膏销售量预计达到 8.71 百万支。

回归分析的主要优点是通过建立一个线性回归模型及一个非线性回归模型，比较两个模型的预测效果找到了更贴近实际且比较方便使用的模型；回归分析的主要缺点是没有充分考虑回归系数的置信区间及因变量的预测区间。

关键词：销售量；价格差；相关分析；相关系数；回归模型；预测；Excel

一、问题的重述

产品的销售量与销售价格、广告投入费用等因素密切相关，表 1 是某公司销售部人员搜集的最近 30 个销售周期（每个周期为四周）的本公司牙膏销售量、销售价格、广告投入费用及同期其他公司同类牙膏的平均销售价格数据。根据这些数据解决以下问题。

（1）合理利用销售量与销售价格、广告投入费用的数据，分析它们之间的关系并建立销售量与销售价格、广告投入费用之间的数学模型。

（2）预测销售量。试预测当价格差为 0.38 元，投入 6.6 百万元的广告费用时该公司牙膏的销售量。

表 1　牙膏销售量与销售价格、广告投入费用的数据

销售周期	公司售价/元	其他厂家均价/元	广告费/百万元	销售量/百万支
1	3.85	3.80	5.50	7.38
2	3.75	4.00	6.75	8.51
3	3.70	4.30	7.25	9.52
4	3.70	3.70	5.50	7.50
5	3.60	3.85	7.00	9.33
6	3.60	3.80	6.50	8.28
7	3.60	3.75	6.75	8.75
8	3.80	3.85	5.25	7.87
9	3.80	3.65	5.25	7.10
10	3.85	4.00	6.00	8.00
11	3.90	4.10	6.50	7.89
12	3.90	4.00	6.25	8.15
13	3.70	4.10	7.00	9.10
14	3.75	4.20	6.90	8.86
15	3.75	4.10	6.80	8.90
16	3.80	4.10	6.80	8.87
17	3.70	4.20	7.10	9.26
18	3.80	4.30	7.00	9.00
19	3.70	4.10	6.80	8.75
20	3.80	3.75	6.50	7.95
21	3.80	3.75	6.25	7.65

续表

销售周期	公司售价/元	其他厂家均价/元	广告费/百万元	销售量/百万支
22	3.75	3.65	6.00	7.27
23	3.70	3.90	6.50	8.00
24	3.55	4.10	7.00	8.50
25	3.60	4.00	6.80	8.75
26	3.65	4.25	6.80	9.21
27	3.70	3.65	6.50	8.27
28	3.75	3.75	5.75	7.67
29	3.80	3.85	5.80	7.93
30	3.70	4.25	6.80	9.26

资料来源：李伯德，王国兴. 经济管理数学模型[M]. 北京：科学出版社，2015.

二、问题的分析

在忽略品牌效应的前提下，顾客更关注的是同类牙膏之间的价格差，而不单单是价格本身，因为牙膏是生活必需品，无论它的价格如何，人们总会需要。因此，研究销售量与销售价格、广告投入费用之间的关系时用其他公司同类牙膏的平均销售价格与本公司的牙膏销售价格之差更为合理。

预测的目的是预测销售量，所以建立以销售量为因变量，以价格差和广告费用为自变量的多元回归模型，即可解决问题1。回归模型通过检验后用所建立的回归模型进行销售量的回归预测即可解决问题2。

三、问题的假设

(1) 假设购买同级别同类产品时不考虑品牌效应。

(2) 以下计算中的数字结果都为近似值，小数点后面统一保留两位有效数字。

四、符号说明

Q：销售量(百万支)。

p_1：价格差($p_1=p_3-p_4$)(元)。

p_2：广告费(百万元)。

p_3：其他公司同类牙膏的平均销售价格(元)。

p_4：本公司销售价格(元)。

r：复相关系数。

r_{Qp_1}：销售量与价格差的相关系数。

r_{Qp_2}：销售量与广告费的相关系数。

α：显著性水平。

n：样本容量。

m：自变量的个数。

$r_{\alpha(n-m-1)}$：相关系数临界值。

五、模型的建立与求解

5.1 数据的处理

首先，需要审核已给的数据，通过审核发现没有异常数据。

其次，根据回归分析的需要，得到其他公司同类牙膏的平均销售价格与本公司的牙膏销售价格差，即

$$p_1 = p_3 - p_4$$

这个过程用 Excel 的算术功能完成即可，得到如表 2 所示的数据。

表 2　带有价格差的牙膏数据

销售周期	p_4/元	p_3/元	p_2/百万元	p_1/元	Q/百万支
1	3.85	3.80	5.50	−0.05	7.38
2	3.75	4.00	6.75	0.25	8.51
3	3.70	4.30	7.25	0.60	9.52
4	3.70	3.70	5.50	0.00	7.50
5	3.60	3.85	7.00	0.25	9.33
6	3.60	3.80	6.50	0.20	8.28
7	3.60	3.75	6.75	0.15	8.75
8	3.80	3.85	5.25	0.05	7.87
9	3.80	3.65	5.25	−0.15	7.10
10	3.85	4.00	6.00	0.15	8.00
11	3.90	4.10	6.50	0.20	7.89
12	3.90	4.00	6.25	0.10	8.15
13	3.70	4.10	7.00	0.40	9.10
14	3.75	4.20	6.90	0.45	8.86
15	3.75	4.10	6.80	0.35	8.90
16	3.80	4.10	6.80	0.30	8.87
17	3.70	4.20	7.10	0.50	9.26
18	3.80	4.30	7.00	0.50	9.00
19	3.70	4.10	6.80	0.40	8.75
20	3.80	3.75	6.50	−0.05	7.95
21	3.80	3.75	6.25	−0.05	7.65
22	3.75	3.65	6.00	−0.10	7.27
23	3.70	3.90	6.50	0.20	8.00
24	3.55	4.10	7.00	0.10	8.50
25	3.60	4.00	6.80	0.50	8.75
26	3.65	4.25	6.80	0.60	9.21
27	3.70	3.65	6.50	−0.05	8.27
28	3.75	3.75	5.75	0.00	7.67
29	3.80	3.85	5.80	0.05	7.93
30	3.70	4.25	6.80	0.55	9.26

5.2 相关分析

回归分析的前提条件是变量之间必须存在高度相关关系。因此,在回归分析之前必须先进行相关分析。相关分析可按照制作散点图、计算相关系数和相关系数的显著性检验三个步骤进行。

1. 制作散点图

为了直观分析变量之间的相关关系,首先制作散点图观察分布情况。可得到如图 1、图 2 所示的散点图,散点图的制作利用 Excel 的图表功能完成。

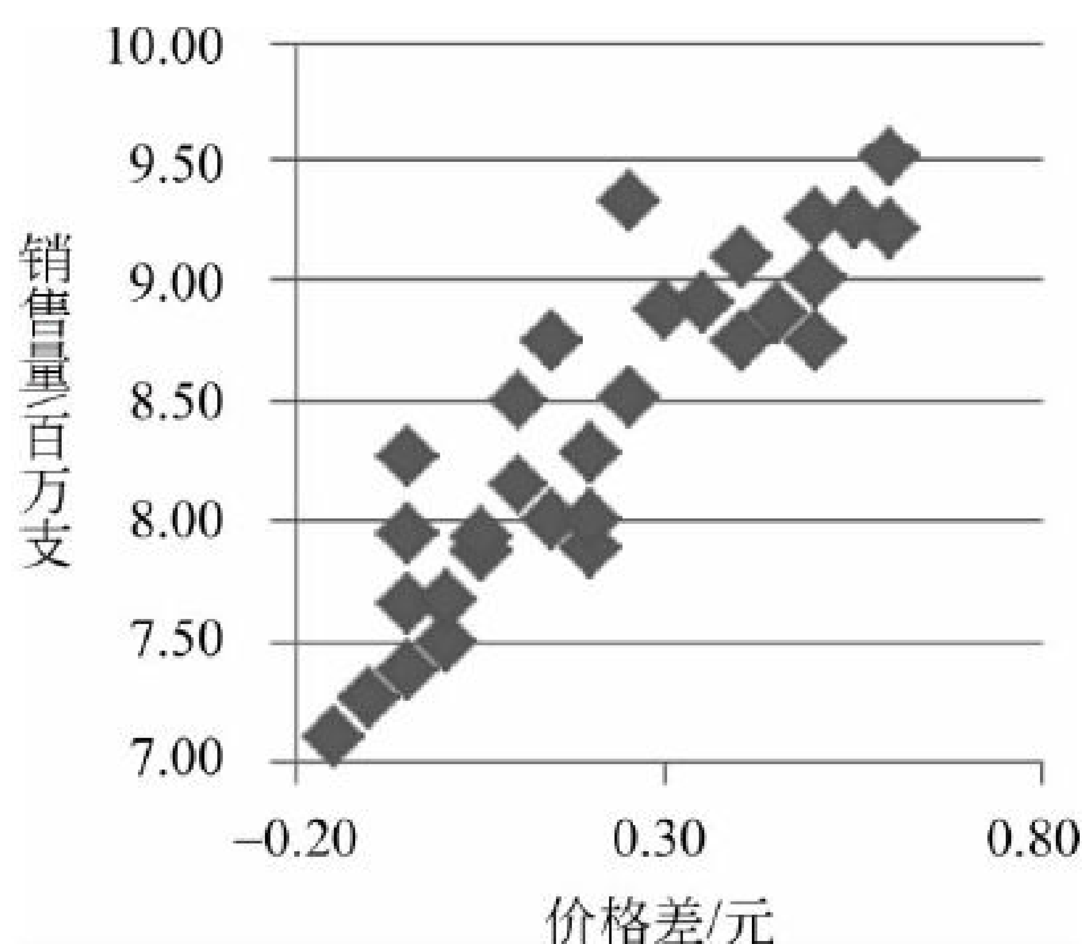


图 1 价格差和销售量的散点图

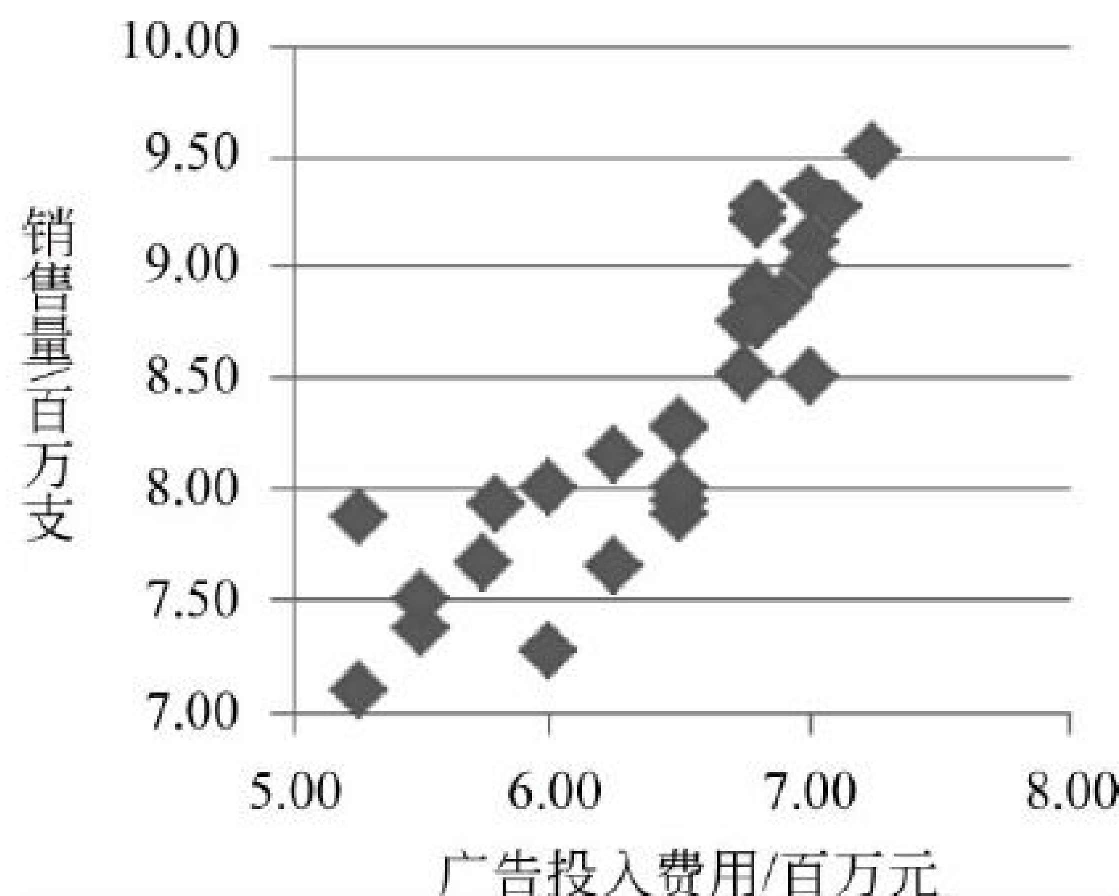


图 2 广告投入费用和销售量的散点图

观察散点图可看出,销售量与价格差、广告投入费用之间存在较好的正相关关系。

2. 计算相关系数

从散点图可以大致看出销售量与价格差、广告投入费用有相关关系,为了衡量他们之间的相关关系强度,接下来计算他们之间的相关系数,利用 Excel 的数据分析工具可得到他们之间的相关系数。具体为,销售量与价格差的相关系数、销售量与广告投入费用的相关系数以及三个变量之间的复相关系数分别为

$$r_{QP_1} = 0.89, \quad r_{QP_2} = 0.86, \quad r = 0.94$$

利用 Excel 计算两个变量的相关系数的具体方法如下。

应用分析工具库可以一次性计算出各变量两两之间的相关系数,称之为相关系数矩阵。具体操作如下。

第一步:单击“数据”菜单,选择“数据分析”项,在“数据分析”项中选择“相关系数”,单击“确定”按钮,此时出现“相关系数”空白对话框,如图 3 和图 4 所示。

第二步:在“输入区域”输入相应数据,用鼠标拖曳选中数据即可,“分组方式”选择逐列,选择“标志位于第一行”,选择输出选项,单击“确定”按钮。此时,即可计算出各项数据两两之间的相关系数矩阵,如图 5 所示。

复相关系数可以通过回归工具得到,具体如图 8 所示。

3. 相关系数的显著性检验

样本相关系数带有一定的随机性,样本容量越小其可信度就越差,因此需要进行显著性检验。

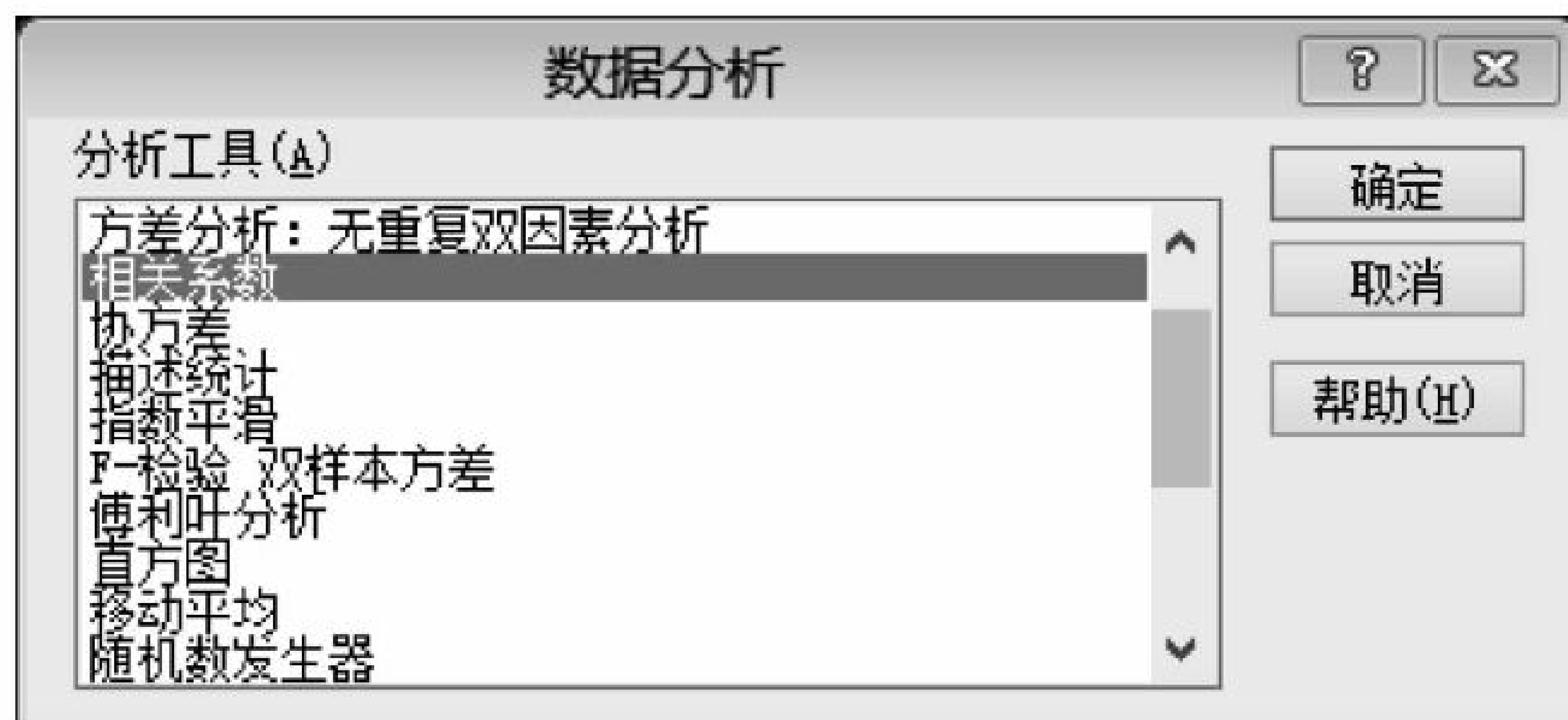


图 3 “数据分析”对话框



图 4 “相关系数”空白对话框

	广告投入费用/百万元	价格差 / 元	销售量/百万支
广告投入费用/百万元	1		
价格差 / 元	0.759 964 027	1	
销售量/百万支	0.875 953 605	0.889 671 764	1

图 5 相关系数矩阵

根据已给数据可知，样本容量 $n=30$；事实上，做相关分析时无须分自变量和因变量，只有做回归分析时需确定因变量和自变量，但在相关分析中，为了检验的需要可以理解为变量当中有一个是因变量，其余的变量为自变量。因此，两个变量的相关系数显著性检验时可认为有一个自变量，即 $m=1$，三个变量之间的显著性检验时可认为有两个自变量，即 $m=2$；显著性水平取 $\alpha=0.05$。

通过查询相关系数检验表可得到相关系数临界值分别为

$$r_{\alpha(n-m-1)}=r_{0.05(28)}=0.36,\quad r_{\alpha(n-m-1)}=r_{30(27)}=0.37$$

因此，

$$r_{Qp_1}>r_{0.05(28)},\quad r_{Qp_2}>r_{0.05(28)},\quad r>r_{30(27)}$$

以上计算出的三个相关系数都大于对应的相关系数临界值，因此相关系数的显著性检验通过。相关系数都大于0.8，因此可确定，变量之间存在高度相关关系，这就说明价格差和广告投入费用对牙膏销售量的影响是显著性的，适合用来做回归分析。

5.3 模型的建立

通过相关分析可知销售量与价格差、广告投入费用存在高度相关关系，三者之间的复相关系数高达 0.94，而根据问题 1 的要求及问题 2 可知其目的是预测销售量，因此可建立以销售量为因变量、以价格差和广告费用为自变量的二元线性回归模型。为了找到更贴近实际的模型，也尝试二元非线性回归模型，比较二者的优劣因素再决定选用哪一种模型。

5.3.1 建立二元线性回归模型（模型Ⅰ）

根据前面的分析可建立如下二元线性回归模型

$$\hat{Q} = \beta_0 + \beta_1 p_1 + \beta_2 p_2 \tag{1}$$

式中：$\hat{Q}$为牙膏销售量的预测值；β_0 为回归截距；β_1、β_2 为回归系数。

确定 β_0、β_1、β_2 的值即可确定模型。确定回归系数的原理是最小二乘法，但计算过程过于烦琐，所以这里利用 Excel 的回归分析工具完成。具体方法如下。

首先，在 Excel 数据中依次单击“数据”→“数据分析”→“回归”→“确定”，如图 6 所示。调出空白“回归”对话框，如图 7 所示。



图 6 “数据分析”窗口



图 7 空白“回归”对话框

其次，在空白“回归”对话框(见图7)中输入因变量销售量和自变量(价格差和广告投入费用)的相应数据、进行相应设置并单击“确定”按钮，即可得到回归分析结果，如图8所示。

SUMMARY OUTPUT								
回归统计								
Multiple R	0.941299347							
R Square	0.886044461							
Adjusted R Square	0.87760331							
标准误差	0.238333761							
观测值	30							
方差分析								
	df	SS	MS	F	Significance F			
回归分析	2	11.92491	5.962453	104.9673	1.84471E-13			
残差	27	1.53368	0.056803					
总计	29	13.45859						
	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
Intercept	4.407493325	0.722329	6.101778	1.62E-06	2.925395832	5.889590818	2.925395832	5.889590818
广告投入费用/百万元	0.563482291	0.119065	4.732556	6.25E-05	0.319180893	0.807783689	0.319180893	0.807783689
价格差/元	1.588285735	0.299432	5.30433	1.35E-05	0.973902127	2.202669342	0.973902127	2.202669342

图8 “回归分析”输出结果

从这个回归分析结果可得到回归截距和回归系数分别为

$$\beta_0 = 4.41,\quad \beta_1 = 1.59,\quad \beta_2 = 0.56$$

代入式(1)即可得到如下二元线性回归模型(模型Ⅰ)：

$$\hat{Q} = 4.41 + 1.59p_1 + 0.56p_2 \tag{2}$$

5.3.2 建立二元非线性回归模型(模型Ⅱ)

为了得到更符合实际的模型并根据数据的特点，下面建立指数回归模型。因自变量的数据中包含零或负数，因此没办法用自变量的倒数形式或对数形式进行线性化，因此采用指数形式的线性化方式，以销售量为因变量 Q，以价格差 p_1 和广告投入费用 p_2 为自变量可得到如下指数回归模型：

$$Q = ae^{b_1p_1+b_2p_2} \tag{3}$$

其中，a、b_1、b_2 为回归系数，为了便于计算回归系数，通过变量代换使式(3)线性化。对式(3)两边取对数，有

$$\ln Q = \ln a + b_1p_1 + b_2p_2 \tag{4}$$

设

$$Q' = \ln Q,\quad p'_1 = p_1,\quad p'_2 = p_2,\quad a' = \ln a$$

代入式(4)可得到如下新的线性化的模型

$$Q' = a' + b_1p'_1 + b_2p'_2 \tag{5}$$

此时，式(5)变成了线性模型，可按照模型Ⅰ中计算回归系数的方法轻松得到回归系数 a、b_1、b_2。按照 $Q'=\ln Q$ 对已给数据进行转换，再利用 Excel 的回归工具可得到如图9所示的回归结果。

从这个回归结果可得到新的回归系数 a'、b_1、b_2 的值，分别为

$$a' = 1.63,\quad b_1 = 0.19,\quad b_2 = 0.07$$

代入式(5)得到

$$Q' = 1.63 + 0.19p'_1 + 0.07p'_2 \tag{6}$$

从而有

回归统计								
Multiple *R*	0.9418308							
R Square	0.8870453							
Adjusted *R* Square	0.8786783							
标准误差	0.0285876							
观测值	30							
方差分析								
	df	SS	MS	*F*	Significance *F*			
回归分析	2	0.173285	0.086643	106.017	1.6376E-13			
残差	27	0.022066	0.000817					
总计	29	0.195351						
	Coefficient	标准误差	*t* Stat	*P*-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
Intercept	1.6307827	0.086642	18.82211	4.72E-17	1.453008282	1.808557	1.4530083	1.808557
广告投入费用/百万元	0.0701134	0.014282	4.909349	3.89E-05	0.04080995	0.099417	0.04081	0.099417
价格差/元	0.1860223	0.035916	5.179337	1.88E-05	0.112328257	0.259716	0.1123283	0.259716

图 9　线性化之后的回归结果

$$\ln Q = 1.63 + 0.19 p_1 + 0.07 p_2 \tag{7}$$

因此可以得到如下指数回归模型(模型Ⅱ)：

$$\hat{Q} = 5.1 e^{0.19 p_1 + 0.07 p_2} \tag{8}$$

5.4　模型的检验

回归模型需从现实意义检验、拟合优度的评价、回归的显著性检验三个方面对其进行检验，检验通过方可以用来预测。上面得到的两个模型进行检验基本类似，下面以模型Ⅰ为例进行详细的检验。

首先，现实意义检验，检验两个回归系数的经济意义是否符合现实。从 $\beta_1 = 1.59$、$\beta_2 = 0.56$ 可知，当价格差($p_3 - p_4$)增加 1 元，即本公司的销售价格降低 1 元或其他公司的平均价格增加 1 元时，本公司的牙膏销售量将增加 1.59 百万支；当广告投入费用增加 1 百万元时，牙膏的销售量将增加 0.56 百万支。这两点基本符合现实经济意义。

其次，拟合优度的评价，评价拟合的好坏程度。拟合程度的好坏主要用决定系数来判断，从回归结果图 8 可知，模型Ⅰ的决定系数为

$$R^2 = 0.89$$

这个数字非常接近 1，说明模型的拟合程度很好。

最后，回归的显著性检验，包括两个方面：①对回归系数的显著性检验；②对整个回归方程的显著性检验。

从回归结果图 8 可知，回归系数 β_0、β_1、β_2 的 P 值分别为 1.62E−06、1.35E−05、6.25E−05(科学记数法)，都远远小于显著性水平 $\alpha = 0.05$，这说明回归系数的影响是显著的，即价格差和广告投入费用对牙膏销售量有显著的影响。

从回归结果图 8 还可知，回归方程的 F 显著性水平为 1.84E−13，该数字同样远远小于显著性水平 $\alpha = 0.05$，这说明回归方程也通过显著性检验。

5.5　问题 2 的求解

从以上三方面的检验可确定模型Ⅰ检验通过。用同样的方法，利用回归结果图 9 中的相应数据确定模型Ⅱ通过检验。因此，可以用模型Ⅰ和模型Ⅱ进行牙膏销售量的预测，但需

要注意的是，用不同的模型预测出的结果有一定的差距。到底选用哪个模型需要综合考虑模型建立过程的合理性与预测的误差等因素。根据问题 2 可知

$$p_1 = 0.38, \quad p_2 = 6.6$$

代入模型Ⅰ可预测出此时的销售量为

$$\hat{Q}_1 = 8.71(\text{百万支})$$

代入模型Ⅱ可预测出此时的销售量为

$$\hat{Q}_2 = 8.70(\text{百万支})$$

$$\hat{Q}_1 - \hat{Q}_2 = 8.71 - 8.70 = 0.01(\text{百万支}) \tag{9}$$

$$\frac{0.01}{8.71} \approx 0.001\,1 \tag{10}$$

说明两个模型的预测值相差 10 000 支，相对差距只有 0.001 1。

当然，仅一个数据不足以说明两个预测值之间的差距，因此，利用模型Ⅰ和模型Ⅱ拟合表 1 中已给的观测值，并做折线图，可得到图 10。

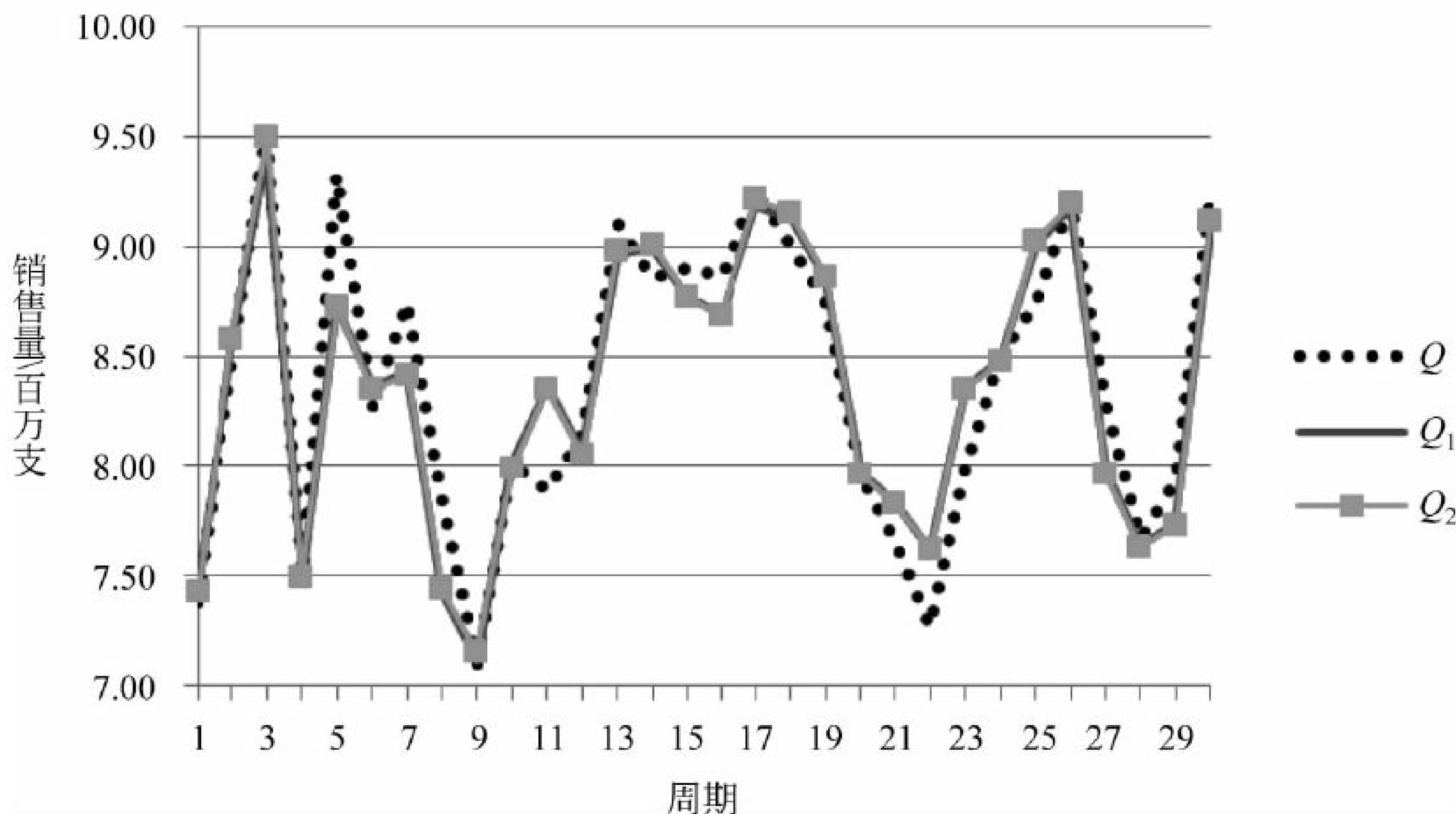


图 10　销售量的观测值及预测值的折线图

图 10 的说明：Q 为观测值，Q_1 为模型Ⅰ的预测值，Q_2 为模型Ⅱ的预测值。

从折线图（见图 10）可看出，两个模型预测值的折线基本重合，且与观测值拟合得非常好，这说明预测误差非常小。

六、模型的分析与评价

6.1　结果的分析

综合考虑式（9）、式（10）和图 10，可以判断出模型Ⅰ和模型Ⅱ的预测结果基本相似，但从模型建立的过程看，模型Ⅱ的建立更为复杂。因此，最终采用模型Ⅰ解决本文中的两个问题。

问题 1：销售量与销售价格差、广告投入费用的关系为高度正线性相关关系。具体数量关系为

$$\hat{Q} = 4.41 + 1.59p_1 + 0.56p_2$$

问题 2：当价格差为 0.38 元，投入 6.6 百万元的广告费用时，该公司牙膏销售量预计达到 8.71 百万支。

6.2 模型的评价

1. 优点

首先，应用题目所给数据之前进行了认真的筛查与审核，以保证数据的准确性，只有保证了数据的准确性，才能保证模型的准确性，从而保证预测误差更小。

其次，回归模型的建立严格按照回归分析的规范步骤来进行，整个过程比较完整。最终的模型没有只靠建立一个模型来决定，而是通过建立两个模型对比优劣再做出决定。

最后，文中除了有数字结论，更是结合散点图、折线图等图表形象直观地表现了部分结论。另外，图表的制作、回归模型的确定及检验都利用 Excel 完成，而 Excel 对于大家来说非常熟悉，用起来比较简单，通俗易懂。

2. 缺点

牙膏销售量的预测值都只给了一个预测值，而现实往往不会太绝对；预测误差的估计不够精细。

3. 模型的改进方向

可进一步考虑回归系数的置信区间及因变量的预测区间，而不单单只是一个值；另外，模型的建立可以考虑把两个自变量（价格差及广告投入费用）对销售量的交叉作用加入进去。

4. 模型的推广

本文只是以牙膏数据为例建立了两个模型，事实上，这些模型可以推广到洗发液、洗衣液、洗衣粉等日化用品以及其他生活必需品的同类问题上。

参考文献

[1] 李伯德，王国兴. 经济管理数学模型[M]. 北京：科学出版社，2015.

[2] 韩中庚. 数学建模竞赛——获奖论文精选与点评[M]. 北京：科学出版社，2013.

教学方法和教学策略

教学方法	讲授与探究式教学法		
师生角色	学生为主体；教师是启发者和辅导者		
实施重点	任务驱动，让学生在实践中探究学习		
实施步骤	教学过程与内容	教学方法与策略	学时
1. 准备	(1) 布置学习任务 (2) 引导学生搜集并预习回归分析的相关知识	引导教学法	1

续表

实施步骤	教学过程与内容	教学方法与策略	学时
2. 实施	(1) 讲授回归分析相关理论 (2) 介绍用 Excel 做回归分析的方法 (3) 分发单项任务完成实际操作 (4) 分发综合任务完成实际操作 (5) 综合分析实际操作结果，完成回归分析任务并形成综合报告	讲授法 思维导图 小组展示	5
3. 检查	(1) 对照任务要求，检查回归分析过程与结果 (2) 发现存在的问题与不足，并做出相应改进	问题引导法 对比分析法	1
4. 评价	(1) 师生共同评价任务完成的效果和质量 (2) 针对不足提出改进措施与注意事项	实操展示 小组讨论	1

成果展示要求

展示成果	1. 财政收入预测问题(学习任务页-zh1)论文(Word 排版) 2. 财政收入预测模型建立过程介绍(PPT)	
展示方式	1. 每组把预测方案制作成幻灯片进行展示 2. 现场操作展示：利用 Excel 做回归分析部分	
展示顺序	抽签决定	
展示时间	每个小组展示时间 7 分钟，回答问题时间 3 分钟	
展示要求	财政收入预测问题论文	• 格式正确，排版规范 • 结构完整，有较强的逻辑性和系统性 • 重点突出，包含重要信息和关联点 • 计算准确，论述周详 • 假设合理，推导严密，结论理由充分 • 条理清楚，行文流畅
	PPT 展示	• 格式和视觉效果较好 • 内容完整，条理清楚，逻辑严谨 • 内容、语言和媒体的选择相辅相成 • 体现团队精神 • 语言富有表现力，易于理解 • 有眼神交流，表情生动 • 语速、语调适中

学习评价——个人工作过程报告

学生信息：______________________________

请在下表中认为描述相符的数字位置画上“√”。

1	小组工作中，个人感觉小组氛围如何？					
	非常好	1	2	3	4	5	非常不好
2	小组成员对本人接受度如何？					
	完全接受	1	2	3	4	5	完全排斥
3	本人的行为是否赢得了小组的理解？					
	非常强	1	2	3	4	5	根本没有
4	在项目中是否学到了东西？					
	非常多	1	2	3	4	5	根本没有
5	对小组工作结果是否满意？					
	非常满意	1	2	3	4	5	完全不满意
6	感觉老师对我们小组工作					
	非常有帮助	1	2	3	4	5	没有帮助和引导

通过下面的调查表，对小组中个人的工作进行评价。

在我的团队中，我能……	评　　价
专业能力 • 按照专业和实际情况展示学习和工作成果 • 独立或在小组中按照时间要求设计工作步骤 • 充分利用专业信息来源和工作辅助工具 • 想出新创意，给出创造性建议 • 区别重要信息和非重要信息	
方法能力 • 获取、阅读、分析、理解信息资料 • 策划工作步骤，在规定时间内实施 • 根据目标，采用基本的工作技巧 • 组织新观点，对比不同意见和评价 • 采用适当的方法展示自己的观点、想法和结果	
社会能力 • 接受并遵守小组内达成一致的交流和对话准则 • 为自己的观点提出理由和依据 • 能够对不同意见做出适当的反应 • 接受小组分配的任务，并参与任务分配工作 • 正确面对小组内的冲突，并找出解决方法 • 向其他学生提供帮助，向其他学生寻求帮助	
自我认识能力 • 自己检查工作成果是否合理、是否正确 • 自己认识到当前工作的步骤及缺点 • 了解个人在小组中的位置和贡献 • 自己确立合适的工作和行动目标	

学习评价——小组工作评价

每个小组中应该确定一名小组长，作为此次工作的负责人。每名组员应该在各项任务中轮流担任小组长。小组长要注意，是否考虑到工作中的每一步，并且实施。当小组工作偏离或背离主题时，应该为小组找到正确的方向。

请小组长在下表右侧各列中，自上而下，将已经完成的工作点用“√”的形式表示出来。

学习任务编号		1	2	3	4	5	6	7	zh1	zh2
计划阶段	分配任务，弄清任务要求									
	确定工作步骤									
	估计所需时间，制定工作时间表									
实施阶段	顺利完成任务									
	互相帮助，并提供咨询									
	认真、仔细地进行和题目相关的工作									
	定时检查工作进度									
	准时准备好报告									
评价阶段	客观地评价工作结果									
	客观地分析合作情况									
	总结对下次小组工作有用的方法									

对自己和小组其他成员的评价：小组中的每个成员都应在小组工作结束后进行此评价。

是否注意到以下小组规则？	学习任务编号（写出评价分数，5＝极好，0＝极差）								
	1	2	3	4	5	6	7	zh1	zh2
小组顺利地开展工作									
小组成员对工作任务详细讨论，并确定目标									
小组成员互相帮助，并且互相关注									
小组成员听取组员意见并且思考									
小组中的所有成员以同样的工作强度参与工作									
每个组员都记录小组及个人的工作过程									
出现问题时，认真讨论并共同解决									
小组成员专心于主题工作，而没有进行其他活动									
小组讨论中没有出现人身攻击或者谩骂的现象									

第五部分

优化模型

学习情境——媒体广告的最优组合

某公司是一家小型公司,它最近要将一种新的旅游产品投放到北京地区的市场中,并且希望通过各种媒体对此产品进行宣传。因此,该公司与当地一家专门从事地区宣传的广告代理商进行了联系,并将此任务完全托付给对方,预算金额总计 250 000 元。

此广告代理商对北京市场有充分的了解,推荐了 6 种不同的媒体,建议进行为期两个月的宣传;对于每种媒体,广告代理商都了解在其上做广告的成本,以及该媒体能够影响到的人数。

此外,广告代理商估计了每种媒体的影响指数,并了解到每种媒体的最大使用量。表 5-1 中列出了上述信息。

表 5-1 媒体广告详细信息

编号	媒体类型	影响人数/人	单位费用/元	最大使用量	影响指数
1	《精品购物指南》周报	12 000	1 500	4 个星期	3
2	《旅游》月刊	1 500	8 000	2 个月	7
3	《都市快报》周刊	2 000	12 000	8 个星期	8
4	北京城市广播	6 000	9 000	60 次广播	2
5	户外广告牌	3 000	24 000	4 个广告牌	6
6	北京电视台生活频道	9 000	51 000	8 次播放	9

该公司希望此广告宣传计划至少能影响到 100 000 人。应该如何选择媒体组合,才能使广告的总影响指数最高?上述理论计算值和实际广告投放有一定的差距。你认为主要的差距有哪些?模型应如何改进?

能力矩阵——优化模型

学生活动	教师控制/教师组织					
教师角色	讲授、辅助咨询			教练		
学习步骤	能力水平					
	认知		应用		创新	
	A1	A2	B1	B2	C1	C2
	识记	理解	应用	分析	创造	评价
1. 优化模型概述	能记住什么是优化模型	能理解优化问题三要素	能判断一个实际问题是否属于优化问题范畴	能分析优化问题难点是什么		
2. 线性规划模型	能记住线性规划模型的概念	能理解解决线性规划模型的步骤;能解决初等线性规划问题	能完成线性规划全部建模过程及文字表述	能从实际问题中进行提炼分析,转化成优化问题	能通过适度简化,创造性地解决实际问题	能对结果进行灵敏性分析
3. 整数规划模型	能记住整数规划模型的概念	能理解整数规划的要求;能解决初等整数规划问题	能完成整数规划全部建模过程及文字表述	能给出整数规划问题的合理解释	能针对实际问题,创造性地提炼、解决	能对结果进行合理化评价
4. 0—1 规划模型	能记住 0—1 规划模型的概念	能理解 0—1 变量的含义;能解决初等 0—1 规划问题	能完成 0—1 规划全部建模过程及文字表述	能给出 0—1 规划问题的合理解释	能针对实际问题,创造性地提炼、解决	
5. 多目标规划模型	能记住多目标规划模型的概念	能理解多目标规划的多种解决方法	能解决初等多目标规划问题	能针对多目标规划问题结果进行合理分析		
6. LINGO 软件求解	能记住 LINGO 软件的初级语法	能理解如何应用 LINGO 软件求解简单优化问题	能应用 LINGO 软件求解整数规划和 0—1 规划模型	能分析 LINGO 软件的求解器状态结果报告和错误提示	能创造性地完成基于集合语言的 LINGO 程序	能对优化模型进行灵敏性分析,并解释释

学习步骤计划书

姓　　名		班　　级	
学习日期		汇报日期	
学习任务			

学习目标	
(1) 我的目标是什么? **注意**:每一个目标都应该以"我能……"开始,并将每一个目标编号;请注明通过哪一个学习任务达到该学习目标。 (2) 我想要学会什么? **注意**:要给出在哪个时间点达到这个目标。	
学习途径	
(1) 我通过哪些途径获得该目标相关信息? (2) 我需要什么材料? (3) 我可以通过哪些方式完成该目标? (4) 达到此目标的过程中会出现什么障碍? (5) 在完成此目标的每一天中,我要完成什么?	
学习证明	
(1) 通过何种展示方式证明我达到这个目标? 如论文写作、PPT 展示、流程图绘制等。 (2) 怎样检测我是否达到这个目标? (3) 请给出预计展示成果的时间。	
学习评价	
(1) 对于自己的学习结果,我是否满意? (2) 我通过什么方法评价自己的学习质量? (3) 我的学习信心是否充足? 我如何克服上面提到的障碍? (4) 我还能做些什么,以完善学习成果?	

学习任务页

学习任务页-1　优化模型概述

学生信息＿＿＿＿＿　　任务开始日期＿＿＿＿＿　　任务分值　10 分

1-A1　优化模型可以解决什么类型的问题?

1-A2　优化模型的三要素是什么?

1-B1-1　下列问题可否应用优化模型解决,请分别说明。

一家汽车厂生产一辆小、中、大型汽车分别需要钢材 1.5 吨、3 吨、5 吨,花费劳动时间分别为 280 小时、250 小时、400 小时,获得利润分别为 2 万元、3 万元、4 万元。厂家每月购得钢材 600 吨,安排劳动时间 60 000 小时。试制定月生产计划,使得工厂的利润最大化。

1-B1-2　下列问题可否应用优化模型解决，请分别说明。

某工厂在计划期内要安排生产甲、乙两种产品。已知生产单位产品所需设备台时及A、B两种原材料的消耗，如表5-2所示。

表5-2　甲、乙两种产品的材料耗费数据

设备及材料	甲	乙	耗费量
设备	1	2	8台时
原料A	4	0	16千克
原料B	0	4	12千克

工厂每生产一件甲产品，可以获利2元；每生产一件乙产品，可以获利3元。试问应该如何安排计划，使该工厂获利最多？

1-B1-3　下列问题可否应用优化模型解决，请分别说明。

一奶制品加工厂用牛奶生产A_1、A_2两种奶制品，1桶牛奶可以在甲车间用12小时加工成3公斤A_1，或者在乙车间用8小时加工成4公斤A_2。根据市场需求，生产的A_1、A_2全部能售出，且每公斤A_1获利24元，每公斤A_2获利16元。现在加工厂每天能得到50桶牛奶的供应，每天正式工人总的劳动时间为480小时，并且甲车间每天至多能加工100公斤A_1，乙车间的加工能力没有限制。试为该厂制订生产计划，使每天获利最大。

1-B2　优化模型的主要方法和原理是什么？



学习任务页-2　线性规划模型

学生信息＿＿＿＿＿＿　　　任务开始日期＿＿＿＿＿＿　　　任务分值＿＿20 分＿＿

2-A1　线性规划建模指的是什么方法？请举实例说明。

2-A2-1　试完成“任务 1-B1-1”的建模过程。

2-A2-2　试完成“任务 1-B1-2”的建模过程。

2-A2-3　试完成“任务 1-B1-3”的建模过程。

2-B1-1　试完成“任务 1-B1-1”模型的文字表述。

2-B1-2　试完成“任务 1-B1-2”模型的文字表述。

2-B1-3　试完成“任务 1-B1-3”模型的文字表述。

2-B2　妈妈希望儿子利用计算机帮助自己进行每年的家庭理财规划。她列出每个月的支出和收入：每月生活费 5 500 元，房租 6 300 元；电话费每 2 个月 1 350 元，燃气、电费和上网费每 6 个月 8 500 元，汽车每个月 3 400 元；纳税每 4 个月 1 000 元；收入每个月 19 000 元工资，加上 1 500 元的补贴。妈妈每个月至少要花费 1 650 元用于各种娱乐，她还希望能有更多的娱乐。那么，应该怎样安排每年的预算，才能使用于娱乐的预算最多？

2-C1-1　某部门现有资金 100 万元，在今后五年内考虑对以下四个项目进行投资。已知项目 1：从第 1 年到第 4 年，每年年初需要投资，并于次年末收回本利 112%；项目 2：第 3 年年初需要投资，到第 5 年年末能收回本利 118%，但规定最大投资额不超过 40 万元；项目 3：第 2 年年初需要投资，到第 5 年年末能收回本利 126%，但规定最多投资额不超过 30 万

元;项目 4:五年内每年年初可购买公债,于当年末归还,并加利息 5%。试确定投资方案,使得收益最大。

2-C1-2　2001 年数学建模竞赛题 C 题的简化变形:基金的优化使用

某校基金会得到一笔数额为 M 元的基金,打算将这笔钱存入银行,在 n 年内每年用部分本息奖励优秀师生,要求各年的奖金额相同,且在 n 年末仍保留原基金数额。现在银行存款的利率如表 5-3 所示。

表 5-3　银行存款的利率表

存期	1 年	2 年	3 年	5 年
税后年利率/%	1.800	2.160	2.592	2.880

校基金会希望获得最佳的基金使用计划,以提高每年的奖金额。请在 $M=5\ 000$ 万元,$n=5$ 年的情况下,设计具体的存款方案。

2-C2　深入分析“任务 1-B1-3”,进一步讨论以下 3 个附加问题。

(1) 若用 35 元可以买到 1 桶牛奶,应否做这项投资?若投资,每天最多购买多少桶牛奶?

(2) 若聘用临时工人来增加劳动时间,付给临时工人的工资最多是每小时多少元?

(3) 由于市场需求变化,每公斤 A1 的获利增加到 30 元,应否应改变生产计划?



学习任务页-3　整数规划模型

学生信息＿＿＿＿＿　　任务开始日期＿＿＿＿＿　　任务分值＿20 分＿

3-A1　整数规划指的是什么？请举实例说明。

3-A2　对于“任务 1-B1-1”，如果要求生产汽车的数量为整数，该如何解决？

3-B1　对于“任务 3-A2”，完成文字表述。

3-B2　某公司是一家小型公司，最近要将一种新的旅游产品投放到北京地区的市场中，并且希望通过各种媒体对此产品进行宣传。因此，该公司与当地一家专门从事地区宣传的广告代理商进行了联系，并将此任务完全托付给对方，预算金额总额为 250 000 元。此广告代理商对北京市场有充分的了解，推荐了 6 种不同的媒体，建议进行为期两个月的宣传；对于每种媒体，广告代理商都了解在其上做广告的成本，以及该媒体能够影响到的人数。此外，广告代理商还估计了每种媒体的影响指数，并了解到每种媒体的最大使用量。表 5-4 中列出了上述信息。

表 5-4　媒体广告详细信息(选择媒体组合)

编号	媒 体 类 型	影响人数/人	单位费用/元	最大使用量	影响指数
1	《精品购物指南》周报	12 000	1 500	4 个星期	3
2	《旅游》月刊	1 500	8 000	2 个月	7
3	《都市快报》周刊	2 000	12 000	8 个星期	8
4	北京城市广播	6 000	9 000	60 次广播	2
5	户外广告牌	3 000	24 000	4 个广告牌	6
6	北京电视台生活频道	9 000	51 000	8 次播放	9

该公司希望此广告宣传计划至少能够影响到 100 000 人。应该如何选择媒体组合,才能使广告的总影响指数最高?

3-C1　某公司客服部需要在网络和电话热线上接待顾客,解答疑问等。根据服务数据历史信息统计得出,一周中每天需要不同数目的雇员:周一到周四每天至少需要 5 人,周五需要 8 人,周六和周日需要 10 人。现为了工作安排方便,规定应聘者需连续工作 5 天,然后休息 2 天。试确定聘用方案,使得在满足需要的前提下,聘用总人数最少。

3-C2　对任务“3-B2”及“3-C1”的结果进行分析评价。



学习任务页-4　0—1 规划模型

学生信息＿＿＿＿＿＿　　任务开始日期＿＿＿＿＿＿　　任务分值＿＿20 分＿＿

4-A1　0—1 规划指的是什么？请举实例说明。

4-A2　一个旅行者的背包最多只能装 6 千克物品。现有 4 件物品的重量和价值分别为 2 千克、3 千克、3 千克、4 千克；100 元、120 元、90 元、115 元。应携带哪些物品，使得携带物品的价值最大？

4-B1　混合泳接力赛由蛙泳、蝶泳、自由泳、仰泳组成。如何根据表 5-5 所示 4 位运动员的 4 种游泳竞赛成绩（分数）安排混合泳接力队，以取得最佳成绩？

表 5-5　竞赛成绩

赛项 / 成绩 / 运动员	蛙泳	蝶泳	自由泳	仰泳
甲	99	60	59	73
乙	79	65	93	87
丙	67	93	63	81
丁	56	79	86	76

4-B2　非洲某酋长国的酋长想把自己的三个女儿 A、B、C 嫁出去，现设恰有 X、Y、Z 三个求婚者。每位求婚者对 A、B、C 愿意支付的彩礼数视其喜欢程度的不同而不同，设为 $\begin{pmatrix} 3 & 5 & 26 \\ 27 & 10 & 28 \\ 1 & 4 & 7 \end{pmatrix}$（矩阵中的元素 c_{ij} 为第 i 个求婚者娶第 j 个女儿愿付的彩礼数）。问酋长应当如何嫁女儿，才能获得最多的彩礼？

4-C1　2008 年数学建模竞赛题 D 题的简化变形：NBA 赛程的评价。

NBA 共有 30 支球队，西部联盟、东部联盟各 15 支。按照地理位置，西部分西南、西北和太平洋 3 个区，东部分东南、中部和大西洋 3 个区，每区 5 支球队。在 5 个多月常规赛中，共有 1 230 场赛事，每支球队要进行 82 场比赛。每支球队与同区的每一球队赛 4 场（主、客各 2 场），与不同部的每一球队赛 2 场（主、客各 1 场），与同部不同区的每一球队有赛 4 场和赛 3 场（2 主 1 客或 2 客 1 主）两种情况，每支球队的主、客场数量相同，且同部 3 个区的球队间保持均衡。试根据赛程，找出与同部不同区球队比赛中，选取赛 3 场的球队的方法。



学习任务页-5　多目标规划模型

学生信息＿＿＿＿＿　　任务开始日期＿＿＿＿＿　　任务分值＿15 分＿

5-A1　多目标规划问题指的是什么？请举实例说明。

5-A2　多目标规划的常用解法有哪些？

5-B1　某工厂在一个计划期内生产甲、乙两种产品，各产品都要消耗 A、B、C 三种不同的资源。每件产品对资源的单位消耗、各种资源的限量以及各产品的单位价格、单位利润和所造成的单位污染如表 5-6 所示。假定产品能全部销售出去，问每期怎样安排生产，才能使利润和产值都最大，且造成的污染最小？

表 5-6　甲、乙两种产品相关信息

产品信息	甲	乙	资源限量
资源 A 单位消耗	9	4	240
资源 B 单位消耗	4	5	200
资源 C 单位消耗	3	10	300
单位产品的价格	400	600	
单位产品的利润	70	120	
单位产品的污染	3	2	

5-B2　某企业拟生产 A 和 B 两种产品，其生产投资费用分别为 2 100 元/吨和 4 800 元/吨。A、B 两种产品的利润分别为 3 600 元/吨和 6 500 元/吨。A、B 产品每月的最大生产能力分别为 5 吨和 8 吨；市场对这两种产品总量的需求每月不少于 9 吨。试问该企业应该如何安排生产计划，才既能满足市场需求，又节约投资，而且使生产利润达到最大？

学习任务页-6　LINGO 软件求解

学生信息__________　　任务开始日期__________　　任务分值____15 分____

6-A1　下面的说法哪些不正确?

(1) 程序以“Model:”开始,以“end”结束,绝对不能省略。

(2) 每条语句必须以“;”结束。

(3) 每行可以有多条语句,语句不可以跨行。

(4) 变量名称必须以字母开头,由字母、数字和下划线组成,区分大小写。

(5) 如没有对变量的取值范围作特殊说明,则默认所有决策变量均非负。

(6) 内部函数以“@”开头。

(7) 说明语句以“!”引导,以“;”结束。

(8) 求目标函数的最大值或者最小值,分别用“max=”或者“min=”来表示。

(9) 可以给语句加上标号,例如,[OBJ] max=2 * x1+3 * x2。

(10) 文件存在扩展名为.lg4 的文档中。

6-A2　如何在 LINGO 中求解如下线性规划问题。

$$
\begin{aligned}
&\min \quad 2x_1 + 3x_2 \\
&\text{s. t.} \\
&\qquad x_1 + x_2 \geqslant 350 \\
&\qquad x_1 \geqslant 100 \\
&\qquad 2x_1 + x_2 \leqslant 600 \\
&\qquad x_1, x_2 \geqslant 0
\end{aligned}
$$

6-B1-1　对于“任务 3-A2”的汽车生产计划问题,利用 LINGO 软件编写程序。

6-B1-2　对于“任务 4-B1”的混合泳指派问题,利用 LINGO 软件编写程序。

6-B2-1　对“任务 6-B1-1”结果进行分析说明。

6-B2-2　对“任务 6-B1-2”结果进行分析说明。

6-C1　对于“任务 4-B1”的混合泳指派问题,利用 LINGO 软件集合语言编写程序。

6-C2　利用 LINGO 软件对“任务 2-C2”进行灵敏性分析。

学习任务页-zh1　工作时间调度模型

学生信息＿＿＿＿＿　　任务开始日期＿＿＿＿＿　　任务分值＿25 分＿

根据“任务 3-C1”中的结论，公司共雇用 10 名客服人员。周末时间为公司繁忙时间段，为了使员工更高效地工作，需要为客服人员制定周末工作时间表。

客户服务的一个工作日分为 12 个两小时长的时间段，每个时间段需要的人员都不同。表 5-7 中列出了周末每个时间段的客服人员需求量。

表 5-7　每个时间段的客服人员需求量

编号	时间段	需要客服人数	编号	时间段	需要客服人数
0	00a. m. ～02a. m.	2	6	12p. m. ～02p. m.	4
1	02a. m. ～04a. m.	1	7	02p. m. ～04p. m.	3
2	04a. m. ～06a. m.	1	8	04p. m. ～06p. m.	3
3	06a. m. ～08a. m.	3	9	06p. m. ～08p. m.	4
4	08a. m. ～10a. m.	4	10	08p. m. ～10p. m.	4
5	10a. m. ～12p. m.	4	11	10p. m. ～12a. m.	2

问题 1：每位客服人员每天工作 8 小时，由于工作强度较大，要求在工作 4 个小时后休息 2 个小时。根据表 5-7 中所列的历史需求数据，需要多少客服人员？

问题 2：目前周末有 10 名客服人员当值，是否可以满足需求？如果不可以，需要安排部分客服人员加班。每天加班时间 2 个小时，且紧随在后一个 4 小时工作时段之后，中间不再休息。请给出客服人员工作时间的安排方案，使需要加班的员工人数最少。

问题 3：可否优化排班计划，使得半夜开始上班的人数最少？



学习任务页-zh2　投资组合模型

学生信息__________　　任务开始日期__________　　任务分值___25 分___

某三种股票(A、B、C)12 年(2002—2013 年)的价格(包括分红在内)及每年的增长情况如表 5-8 所示。其中,最后一列给出了相应年份 500 种股票的价格指数增长情况。例如,表中第一个数据 1.300 表示股票 A 在 1999 年年末的价值是其年初价值的 1.300 倍,即收益为 30%。其余数据以此类推。

假设在 2014 年有一笔资金准备投资这三种股票,并期望年收益至少达到 15%。那么,应该如何投资?当期望的年收益率变化时,投资组合和相应的风险如何变化?

表 5-8　股票收益数据

年份	股票 A	股票 B	股票 C	股票指数
2002	1.300	1.225	1.149	1.258 997
2003	1.103	1.290	1.260	1.197 526
2004	1.216	1.216	1.419	1.364 361
2005	0.954	0.728	0.922	0.919 287
2006	0.929	1.144	1.169	1.070 8
2007	1.056	1.107	0.965	1.055 012
2008	1.038	1.321	1.133	1.187 925
2009	1.089	1.305	1.732	1.317 13
2010	1.090	1.195	1.021	1.240 164
2011	1.083	1.390	1.131	1.183 675
2012	1.035	0.928	1.006	0.990 108
2013	1.176	1.175	1.908	1.526 236

学习信息库 1——优化模型基本概念

与最大、最小、最长、最短等有关的问题是最优化问题。

1. 优化问题分类

优化问题可分为：线性规划、非线性规划、整数规划、0—1 规划、(多)目标规划、(与时间有关的)动态规划、(系数是随机变量的)随机规划等。

2. 优化问题三要素

(1) 决策变量通常是某一问题需要求解的未知量，用 n 维向量 $x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 表示。当对 x 赋值后，它通常称为该问题的一个解。

(2) 目标函数，通常是某一问题需要优化(最大或最小)的那个目标的数学表达式。它是决策变量 x 的函数，可以抽象地记作 $f(x)$。

(3) 约束条件，由该问题对决策变量的现实条件给出，即 x 允许的取值范围为 $x\in\Omega$，Ω 称为可行域。常用一组关于 x 的等式 $h_i(x)=0(i=1,2,\cdots,m)$ 和(或)不等式 $g_j(x)\leqslant 0(j=1,2,\cdots,n)$ 来界定，分别称为等式约束和不等式约束。

3. 优化模型

$$\begin{aligned}\text{opt}\quad & z=f(x)\\ \text{s. t.}\quad & h(x)=0\quad (i=1,2,\cdots,m)\\ & g(x)\leqslant 0\quad (j=1,2,\cdots,n)\end{aligned}$$

这里，opt 是最优化的意思，可以是 min(求极小，即 minimum 的缩写)或 max(求极大，即 maximize 的缩写)的两者之一；s. t.(即 subject to)是“受约束于”之意。

4. 线性规划(LP)

(1) 带有线性约束条件的多变量线性函数的最值问题。

(2) 当目标函数和约束条件都是决策变量的线性函数时，称为线性规划问题，否则称为非线性规划问题。

(3) 1939 年，苏联数学家康托洛维奇发表《生产组织与计划中的数学问题》。

(4) 1947 年，美国数学家乔治・丹契克、冯・诺伊曼提出线性规划的一般模型及理论。

5. 线性规划(LP)的一般形式

线性规划(LP)的一般形式为

$$\min f(x)=\sum_{i=1}^{n}c_i x_i\quad (i=1,2,\cdots,n)$$

$$\text{s. t.}\ \sum_{i=1}^{n}a_i x_i\geqslant b_i\quad (\leqslant b_i,\ =b_i)$$

学习信息库 2——优化问题解决方法串讲

1. 投资问题——如何入手解决优化问题

某部门现有资金 100 万元，在今后五年内，考虑对以下四个项目进行投资。已知项目 1：从第 1 年到第 4 年，每年年初需要投资，并于次年末收回本利 112%；项目 2：第 3 年年初需要投资，到第 5 年年末能收回本利 118%，但规定最多投资额不超过 40 万元；项目 3：第 2 年年初需要投资，到第 5 年年末能收回本利 126%，但规定最多投资额不超过 30 万元；项目 4：五年内每年年初可购买公债，于当年年末归还，并加利息 5%。试确定投资方案，使得收益最大。

(1) 提示

① 做优化问题的最关键一点是确定决策变量。也就是，到底要决策什么事情。设身处地地思考问题，会更好地解决问题。

② 什么叫作确定投资方案？什么叫作如何投资？搞清楚优化问题三要素(决策变量、目标函数、约束条件)具体是什么，再翻译成数学量化的语言，问题就迎刃而解。

③ 本题关键点 1：每年是否可以、是否应该将全部 100 万元投资出去？

④ 本题关键点 2：每个项目到底是如何投资的？哪年投资？哪年收回？

⑤ 下面重温建模步骤，如图 5-1 所示。

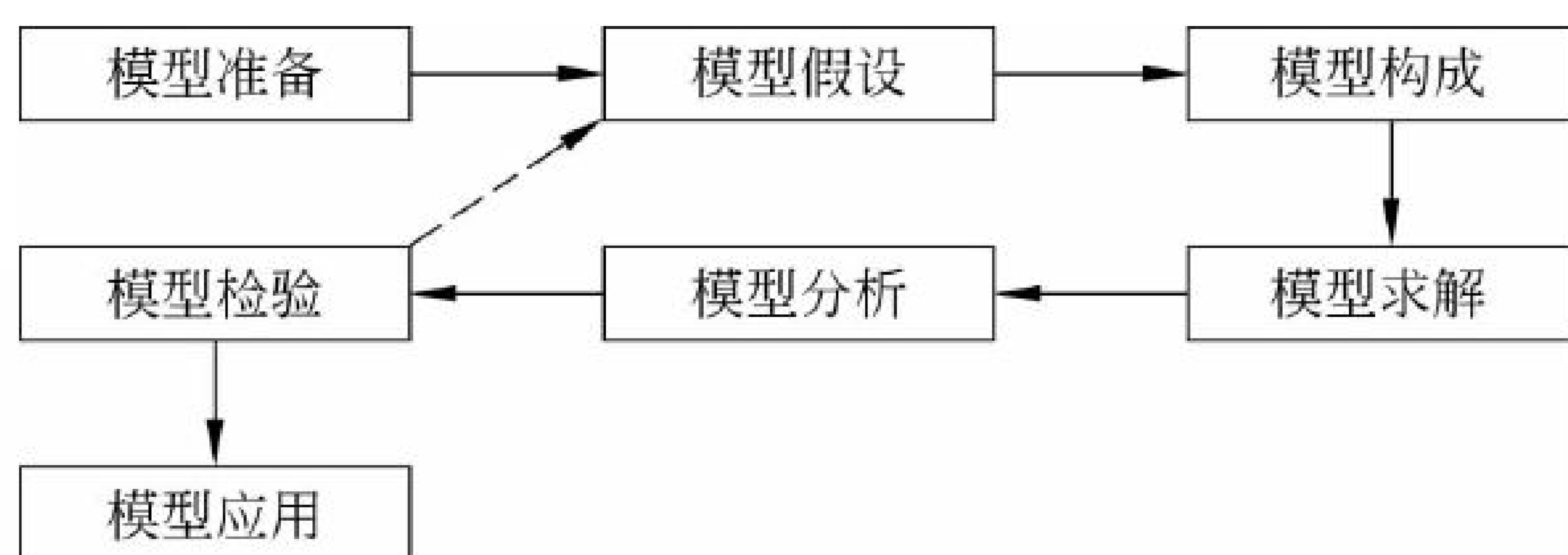


图 5-1　建模步骤

(2) 建立模型

设 x_{ij} 表示第 i 年投资项目 j 的资金金额($i=1,2,3,4,5,j=1,2,3,4$)。

因为项目 4 每年年初可以投资，年末即可收回本利，故每年手头不留资金，全部投出。

以下式(1)～式(5)，等号左侧表示当年可以投资金额总和，右侧表示当年手里资金总额。

第 1 年年初：　$x_{11}+x_{14}=100$　(1)

第 2 年年初：　$x_{21}+x_{23}+x_{24}=(1+5\%)x_{14}$　(2)

第 3 年年初：　$x_{31}+x_{32}+x_{34}=112\%x_{11}+(1+5\%)x_{24}$　(3)

第 4 年年初：　$x_{41}+x_{44}=112\%x_{21}+(1+5\%)x_{34}$　(4)

第 5 年年初：　$x_{54}=112\%x_{31}+(1+5\%)x_{44}$　(5)

由于项目 2 和项目 3 有投资金额限制，所以

$$x_{32}\leqslant 40,\quad x_{23}\leqslant 30$$

此外，还有非负约束条件：$x_{ij}\geqslant 0(i=1,2,3,4,5;j=1,2,3,4)$。

投资目标是第 5 年年末收益最大，使得目标函数

$$Z = 112\% x_{41} + 118\% x_{32} + 126\% x_{23} + (1 + 5\%) x_{54}$$

达到最大。

综上所述，可得优化模型为

$$\max = 112\% x_{41} + 118\% x_{32} + 126\% x_{23} + (1 + 5\%) x_{54}$$

$$\text{s.t.} \begin{cases} x_{11} + x_{14} = 100 \\ x_{21} + x_{23} + x_{24} = (1 + 5\%) x_{14} \\ x_{31} + x_{32} + x_{34} = 112\% x_{11} + (1 + 5\%) x_{24} \\ x_{41} + x_{44} = 112\% x_{21} + (1 + 5\%) x_{34} \\ x_{54} = 112\% x_{31} + (1 + 5\%) x_{44} \\ x_{32} \leqslant 40,\ x_{23} \leqslant 30 \\ x_{ij} \geqslant 0,\ i = 1,2,3,4,5;\ j = 1,2,3,4 \end{cases}$$

(3) LINGO 软件介绍及语法规则

LINGO 软件包由美国芝加哥大学的 Linus Scharge 教授于 1980 年前后开发，专门用于求解最优化问题，后经不断完善和扩充，并成立 LINDO 公司进行商业化运作，取得了巨大的成功。全球《财富》杂志“500 强企业”中，一半以上使用该公司产品，其中前 25 强企业中有 23 家使用该产品。该软件包功能强大，版本很多。免费版本是演示版(试用版)，与正式版在功能上类似，只是能够求解问题的规模受到限制，总变量数不超过 30 个。LINGO 是英文 Linear Interactive and General Optimizer 字首的缩写，即“交互式的线性和通用优化求解器”，它还可以用来求解非线性规划。

LINGO 的最大特色在于允许决策变量是整数，而且执行速度很快；LINGO 实际上还是一种建模语言，即使对优化方面的专业知识了解不多的用户，也能方便地执行输入、求解，并能快速地得到复杂优化问题的高质量的解。

下面是其语法规则。

① 在 LINGO Model 窗口直接按问题模型输入。

② 程序以“Model:”开始，以“end”结束。对于较简单的模型，这两条语句可以省略。

③ 每条语句必须以“;”结束，每行可以有多条语句，语句也可以跨行。

④ 变量名称必须以字母(A～Z)开头，由字母、数字和下划线组成，长度不超过 32 个字符，不区分大小写。

⑤ 如无对变量的取值范围作特殊说明，则默认所有决策变量均非负。

⑥ 表达式中可以有括号，右端也可以有数学符号(但尽量避免)。

⑦ 内部函数以“@”开头。

⑧ 说明语句以“!”引导，以“;”结束。

⑨ 求目标函数的最大值或者最小值，分别用“max=”或者“min=”来表示。

⑩ 可以给语句加上标号，例如，[OBJ]max=2 * x1+3 * x2。

⑪ 文件存在扩展名为.lg4 的文档中。

利用上述 LINGO 软件语法规则写出程序如下所示。

(4) LINGO 程序

```
model:
```

```
max=1.12*x41+1.18*x32+1.26*x23+(1+0.05)*x54;
x11+x14=100;
x21+x23+x24=(1+0.05)*x14;
x31+x32+x34=1.12*x11+(1+0.05)*x24;
x41+x44=1.12*x21+(1+0.05)*x34;
x54=1.12*x31+(1+0.05)*x44;

x32<=40;
x23<=30;
end
```

(5) LINGO 结果及解读

① 求解器状态，如图 5-2 所示。

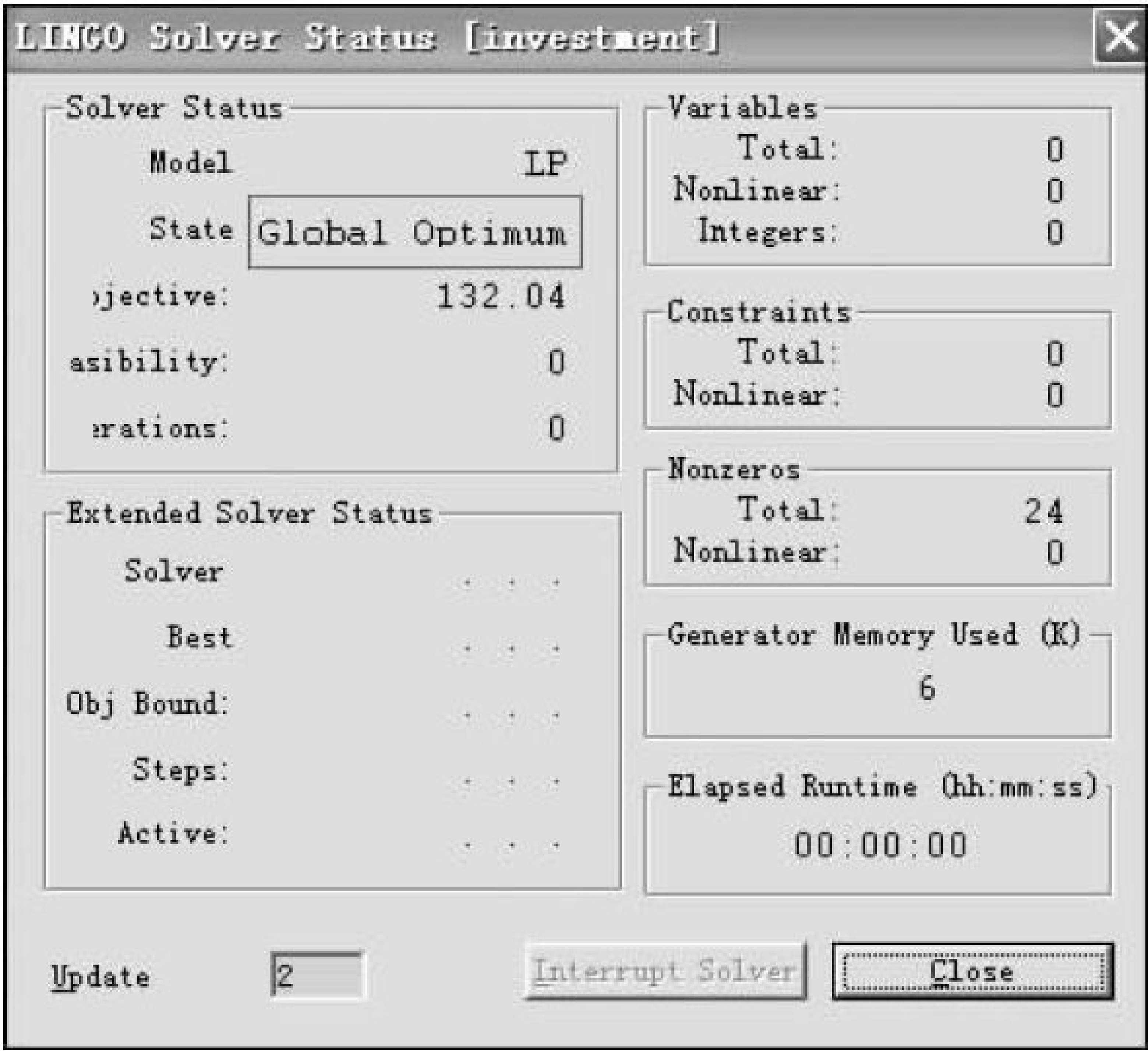


图 5-2　求解器状态

图 5-2 中的 Global Optimum 说明本问题取到全局最优解。

下面就求解器中的其他信息进行解释。

变量数目：变量总数(Total)、非线型变量数(Nonlinear)、整数变量数(Integers)。

约束变量：约束总数(Total)、非线性约束个数(Nonlinear)。

非线性系数数量：总数(Total)、非线性项的系数个数(Nonlinear)。

内存使用量：单位为千字节。

求解花费时间：显示格式"时：分：秒："。

② LINGO 结果，如下所示。

Global optimal solution found at iteration:　　0

Objective value:(最优目标函数值)　　132.0400

Variable	Value	Reduced Cost
变量名称	最优解中的变量值	缩减成本系数

注：Reduced Cost 表示最优单纯形表中判别数所在行的变量的系数。最优解中，变量的 Reduced Cost 值自动为零。

X41	0.000000	0.000000
X32	40.00000	0.000000
X23	30.00000	0.000000
X54	44.80000	0.000000
X11	71.42857	0.000000
X14	28.57143	0.000000
X21	0.000000	0.000000
X24	0.000000	0.1960000E-01
X31	40.00000	0.000000
X34	0.000000	0.000000
X44	0.000000	0.1750000E-01

Row	Slack or Surplus	Dual Price
模型中的行号	松弛变量(松弛或剩余)	对偶价格(影子价格)

注：松弛或剩余，即约束条件左边与右边的差值，对于"＜＝"不等式，右边减左边的差值成为 Slack(松弛)；对于"＞＝"不等式，左边减右边的差值成为 Surplus(剩余)；当约束条件的左、右两边相等时，松弛或剩余的值为零；如果约束条件无法满足，即没有可行解，则松弛或剩余的值为负数。

对偶价格，最优单纯形表中判别数所在行的松弛变量的系数。表示当对应约束条件有微小变动时，目标函数的变化率。例如，若某个变量的对偶价格为 a，表示对应约束条件中，不等式右端如果增加一个单位，目标函数将增加 a 个单位(max 型问题)(若为 min 型问题，为减少)。

1	132.0400	1.000000
2	0.000000	1.317120
3	0.000000	1.254400
4	0.000000	1.176000
5	0.000000	1.120000
6	0.000000	1.050000
7	0.000000	0.4000000E-02
8	0.000000	0.5600000E-02

注：关于 Global optimal solution(全局最优解)。

LINGO 默认的 NLP 算法只进行局部搜索，这可能导致在局部最优解处停止，错过全局最优解。可以通过设置 Global Solver(全局最优求解器)选项卡，选择全局最优算法，如图 5-3 所示。

此时，求解花费的时间可能很长。

模型结果说明如下。

非零变量值为：

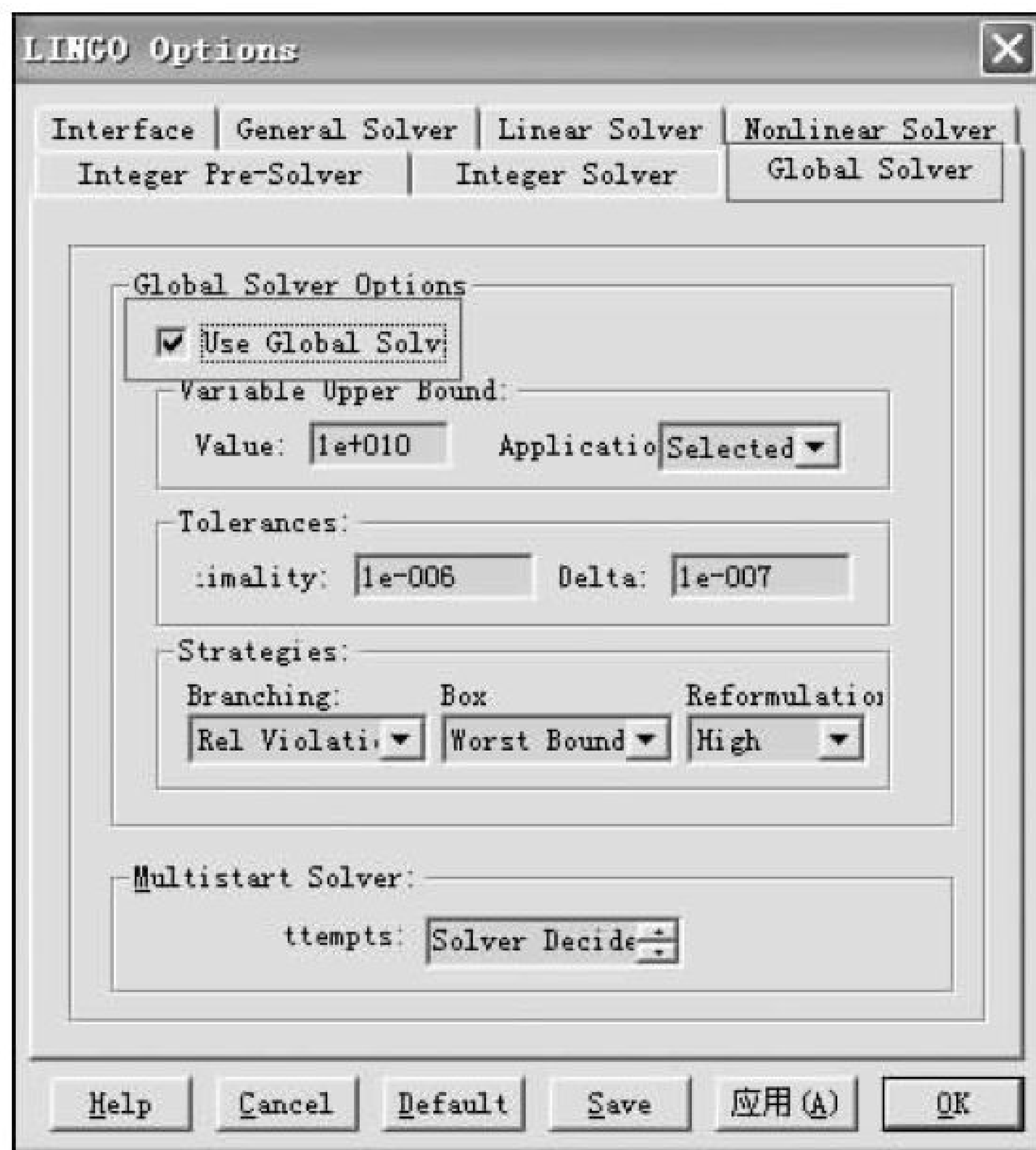


图 5-3　设置 Global Solver 选项卡

```
X32        40.00000
X23        30.00000
X54        44.80000
X11        71.42857
X14        28.57143
X31        40.00000
```

将其翻译成现实语言，即第 1 年投资项目 1 的资金为 72.428 57 万元，剩余资金全部投资在项目 4；第 2 年年初投资项目 3 的资金为 30 万元（从第 1 年项目 4 中收回的）；第 3 年年初投资项目 2 的资金为 40 万元，并投资项目 1 资金 40 万元（这 80 万元是从第 1 年项目 1 中收回的）；第 4 年不投资；第 5 年年初投资项目 4 资金 44.8 万元（从第 3 年项目 1 中收回的）。

最终收益 132.04 万元。

③ 灵敏性分析（Range，Ctrl＋R），如下所述。

当线性规划问题中的常数发生变化（由于测量误差，或具有多个取值可能）时，最优解是否会随之变化？

例如，当某项目的收益发生变化时，投资计划是否应马上随之改变？

可以稍微改变收益率，观察最优解的变化，讨论参数的灵敏性。

用该命令产生当前模型的灵敏性分析报告。

最优解保持不变的情况下，目标函数的系数的变化范围。

在影子价格和缩减成本系数不变的前提下，约束条件右边常数的变化范围。

默认状态下，灵敏度计算是关闭的，运行 LINGO→Options...，选择 General Solver。在 Dual Computations 列表框中，选择 Prices & Ranges 选项，并确定，如图 5-4 所示。

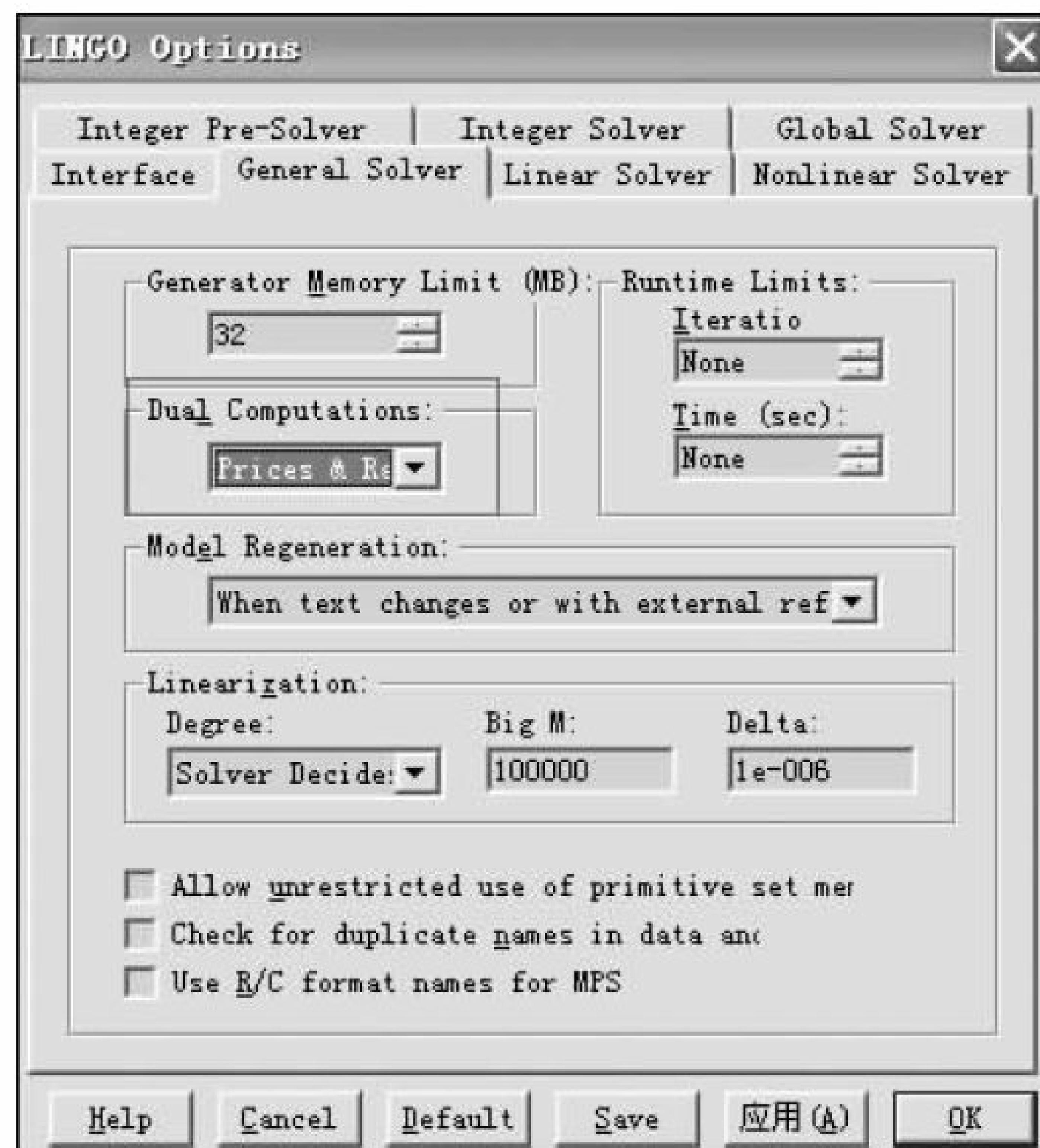


图 5-4　开户灵敏度计算

注意：灵敏性分析耗费相当多的求解时间，因此如无必要，无须激活。

Ranges in which the basis is unchanged:(在哪个范围内最优基不变)

Objective Coefficient Ranges

目标函数系数的变化范围

Variable 变量名称	Current Coefficient 当前系数	Allowable Increase 允许上调的界限	Allowable Decrease 允许下调的界限
X41	1.120000	0.0	0.0
X32	1.180000	0.0	0.0
X23	1.260000	0.0	0.0
X54	1.050000	0.0	0.0
X11	0.0	0.0	0.0
X14	0.0	0.0	0.0
X21	0.0	0.0	0.0
X24	0.0	0.0	0.0
X31	0.0	0.0	0.0
X34	0.0	0.0	0.0
X44	0.0	0.0	0.0

Righthand Side Ranges

约束条件右边常数的变化范围

Row 行标号	Current RHS 当前右边常数值	Allowable Increase	Allowable Decrease

在影子价格和缩减成本系数都不变的条件下，允许上调和下调的界限

```
2        100.0000        0.0        0.0
3        0.0             0.0        0.0
4        0.0             0.0        0.0
5        0.0             0.0        0.0
6        0.0             0.0        0.0
7        40.00000        0.0        0.0
8        30.00000        0.0        0.0
```

2. 2001 年数学建模竞赛题 C 题的简化变形：基金的优化使用

某校基金会得到一笔数额为 M 元的基金，打算将这笔钱存入银行，在 n 年内每年用部分本息奖励优秀师生，要求各年的奖金额相同，且在 n 年年末仍保留原基金数额。现行银行存款的利率如表 5-9 所示。

表 5-9　利率表

存期/年	1	2	3	5
税后年利率/%	1.800	2.160	2.592	2.880

校基金会希望获得最佳的基金使用计划，以提高每年的奖金额。请在 $M=5\ 000$ 万元，$n=5$ 年的情况下，设计存款方案。

(1) 初级模型建立

首先需要做出重要假设：奖金发放时间；银行按单利还是复利计算。

假设 1：奖金于基金到位 1 年后开始发放，以后每隔一年发放一次(为了获得更优的基金使用计划)。

假设 2：银行按单利发放(此假设影响不大，单、复利只是乘以 m，或是 m 次乘幂的差别，m 表示存期)。

设基金总额 $M=5\ 000$ 万元，投资年限 $n=5$。

各年期的利率为 $a_i(i=1,2,3,4)$。

注：5 年期利率按照顺序设为 a_4，好处在于之后建立模型、写 LINGO 程序时，方便引入集合的概念。

设第 i 年年初存第 j 种存期的资金金额为 $x_{ij}\geqslant 0(i=1,2,3,4,5;j=1,2,3,4)$。

由于第 5 年年末要收回所有资金，所以，

```
X24=0;
X34=0;
X43=0; x44=0;
X52=0; x53=0; x54=0;
```

由于可以存 1 年定期，故从第 2 年年初开始，每年年初发放完奖金后，都将手头所有资金都投资出去。故有下列约束条件。

第 1 年年初：

$$\sum_{j=1}^{4} x_{1j} = M \quad (\text{全部基金投资出去})$$

第 2 年年初：

$x_{11}(1+a_1\%) \geqslant \text{award}$ （上一年投资的收益足够支付奖金）

$\sum_{j=1}^{4} x_{2j} = x_{11}(1+a_1\%) - \text{award}$ （发放奖金后剩余的资金全部投资出去）

规律性技巧：x 的下标和为 2，如 x_{11}。

第 3 年年初：

$$x_{12}(1+a_2\times 2\%) + x_{21}(1+a_1\%) \geqslant \text{award}$$

$$\sum_{j=1}^{4} x_{3j} = x_{12}(1+a_2\times 2\%) + x_{21}(1+a_1\%) - \text{award}$$

规律性技巧：x 的下标和为 3，如 x_{12} 和 x_{21}。

第 4 年年初：

$$x_{13}(1+a_3\times 3\%) + x_{22}(1+a_2\times 2\%) + x_{31}(1+a_1\%) \geqslant \text{award}$$

$$\sum_{j=1}^{4} x_{4j} = x_{13}(1+a_3\times 3\%) + x_{22}(1+a_2\times 2\%) + x_{31}(1+a_1\%) - \text{award}$$

规律性技巧：x 的下标和为 4。

第 5 年年初：

$$x_{23}(1+a_3\times 3\%) + x_{32}(1+a_2\times 2\%) + x_{41}(1+a_1\%) \geqslant \text{award}$$

$$\sum_{j=1}^{4} x_{5j} = x_{51} = x_{23}(1+a_3\times 3\%) + x_{32}(1+a_2\times 2\%) + x_{41}(1+a_1\%) - \text{award}$$

规律性技巧：x 的下标和为 5。

第 5 年年底：

$$x_{14}(1+a_4\times 5\%) + x_{33}(1+a_3\times 3\%) + x_{42}(1+a_2\times 2\%) + x_{51}(1+a_1\%)$$
$$= M + \text{award}$$

注：所得 x_{14} 与之前规律性技巧不符合，因为这里设的时候是按照年限的顺序，而实际上是 5 年期。本质上没有冲突。

(2) 模型一的 LINGO 程序及求解

```
model:
sets:
year/1..5/;
cunqi/1..4/:a;
link(year,cunqi):x
endsets

data:
M=5000;
a=1.8 2.16 2.592 2.88;
enddata

max=award;

x(2,4)=0;
```

```
x(3,4)=0;
x(4,3)=0;x(4,4)=0;
x(5,2)=0;x(5,3)=0;x(5,4)=0;

@sum(cunqi(j) : x(1,j))=M;

x(1,1) * (1+a(1)/100) >=award;
@sum(cunqi(j) : x(2,j))=x(1,1) * (1+a(1)/100) -award;

x(1,2) * (1+a(2) * 2/100)+x(2,1) * (1+a(1)/100) >=award;
@sum(cunqi(j) : x(3,j))=x(1,2) * (1+a(2) * 2/100)+x(2,1) * (1+a(1)/100) -award;

x(1,3) * (1+a(3) * 3/100)+x(2,2) * (1+a(2) * 2/100)+x(3,1) * (1+a(1)/100) >=award;
@sum(cunqi(j) : x(4,j))=x(1,3) * (1+a(3) * 3/100)+x(2,2) * (1+a(2) * 2/100)+x(3,1) * (1+a(1)/100) -award;

x(2,3) * (1+a(3) * 3/100)+x(3,2) * (1+a(2) * 2/100)+x(4,1) * (1+a(1)/100) >=award;
@sum(cunqi(j) : x(5,j))=x(2,3) * (1+a(3) * 3/100)+x(3,2) * (1+a(2) * 2/100)+x(4,1) * (1+a(1)/100) -award;

x(1,4) * (1+a(4) * 5/100)+x(3,3) * (1+a(3) * 3/100)+x(4,2) * (1+a(2) * 2/100)+x(5,1) * (1+a(1)/100)=M+
award;

end
```

结果说明：

```
Objective value:          135.2227
非零变量为：X(1, 1)        256.0796
          X(1, 2)        129.6230
          X(1, 3)        125.4664
          X(1, 4)        4488.831
          X(2, 3)        125.4664
```

（3）模型二建立（深入分析，化繁为简）

问题分析：由于要求 n 年末仍保留原基金数额 M，则 n 年中发放的奖金总额为利息。如全部基金都存为 1 年定期，每年均用到期利息发放奖金，则每年的奖金数额为 5 000×1.8%=90(万元)。这是没有优化的最简单的一种基金运作方式。显然，准备 2 年后才用到的奖金应当存成 2 年定期，比存两次 1 年定期更为收益高。由此，需要细致分析各存期的收益情况，特计算各存期的收益比，如表 5-10 所示。

表 5-10 各存期年限对应的最优收益比 $b_i(i=1,2,3,4,5)$

存期/年	1	2	3	4(3+1)	5
最优收益比	1.018	1.043 2	1.077 76	1.097 159 68	1.144

$$\text{收益比 } b=(\text{本金}+\text{利息})/\text{本金}$$

例如，存期 2 年的收益比 b_2 为

$$b_2=1+2.16\%\times 2=1.043\,2。$$

分析表 5-10 可知：

① 一次性存为最长期，优于两个（或两个以上）较短期的组合（即中途转存）。

② 当存期年限需要组合时，收益比与组合的先后次序无关。

决策变量的设定：将基金分块，以满足不同的需求。

最初的基金分配包括分别存为 1～5 年定期的几部分。其中，1～4 年定期到期后，作为当年发放的奖金；而存 5 年定期的，除了需要发放第 5 年的奖金外，还需要剩余基金额与原始金额一致，即为 M。

故将基金总额 M 分成 5+1 份，分别用 $x_1,x_2,\cdots,x_6$ 表示。其中，$x_1,x_2,\cdots,x_5$ 分别存为 1～5 年定期，到期后，本息合计用于当年发放奖金；x_6 表示存 5 年定期，到期的本息合计等于原基金总额 M。

设 award 表示每年用于奖励优秀师生的奖金额，$b_i(i=1,2,3,4,5)$ 表示第 i 年的最优收益比。

本优化问题欲使目标函数每年的奖金额 award 达到最大，即

$$\max = \text{award}$$

约束条件为：

① 各年度奖金额相同，均为 award，即

$$b_i \times x_i = \text{award} \quad (i = 1,2,3,4,5)$$

② n 年末发放奖金后，仍保持原有基金总额 M，即

$$b_5 \times x_6 = M$$

③ 各投资额非负，即 $x_i \geqslant 0$。

（4）模型二的 LINGO 程序及求解

```
model:
sets:
year/1..5/:x,b;
endsets

data:
M=5000;
b=1.018 1.0432 1.07776 1.09715968 1.144;
enddata

max=award;

@for(year : b* x=award;);
@sum(year : x)+x6=M;
b(5) * x6=M;

end
```

结果说明：

```
X(1)                    132.8317
```

```
X(2)          129.6230
X(3)          125.4664
X(4)          123.2479
X(5)          118.2016
X6            4370.629

Objective value:          135.2227
```

注：此最优解和模型一一致，区别在于4年定期的投资方式是：先定期1年，再3年；和先3年，再定期1年是一样的。

(5) 练习(工程进度问题)

某城市在未来五年内将启动4个城市住房改造工程。每项工程有不同的开始时间，工程周期也不同，如表5-11所示。

表 5-11 项目的基本数据

工　程	第1年	第2年	第3年	第4年	第5年	总费用/千万元	年收入/万元
工程1	开始		结束			5	50
工程2		开始			结束	8	70
工程3	开始				结束	15	150
工程4			开始	结束		1.2	20
预算/千万元	3	6	7	7	7		

工程1和工程4必须在规定的周期内全部完成。必要时，其余的两项工程可以在预算的限制内部分完成。然而，每个工程在其规定时间内必须至少完成25%。每年年底，工程完成的部分立刻入住，并且实现一定比例的收入。例如，如果工程1在第1年完成40%，第3年完成剩下的60%，在五年计划范围内的相应收入是0.4×50(第2年)+0.4×50(第3年)+(0.4+0.6)×50(第4年)+(0.4+0.6)×50(第5年)=(0.4×4+0.6×2)×50(单位：万元)。

试为工程确定最优的时间进度表，使得五年内的总收入达到最大。

解　设工程i在第j年完成的进度为x_{ij}，设年收入为in(i)(万元)，总费用为cost(i)(千万元)，预算为yu(j)(千万元)，则

$$\max = \sum_i \left(\mathrm{in}(i) \cdot \sum_j x(i,j) \cdot (5-j)\right)$$

$$\sum_i \mathrm{cost}(i) \cdot x(i,j) \leqslant \mathrm{yu}(j) \quad (j = 1,2,3,4,5)$$

LINGO程序如下所示。

```
model:
sets:
project/1..4/:cost,in;
year/1..5/:yu;
links(project,year):x;
endsets
```

```
data:
cost= 5 8 15 1.2;
in= 50 70 150 20;
yu= 3 6 7 7 7;
enddata

max= @sum(project(i) : in(i) * @sum(year(j):x(i,j) * (5- j)));
@sum(year(j) : x(1,j))= 1; x(1,4)= 0; x(1,5)= 0;
@sum(year(j) : x(2,j))> = 0.25; x(2,1)= 0; @sum(year(j) : x(2,j)) < = 1;
@sum(year(j) : x(3,j))> = 0.25; @sum(year(j) : x(3,j)) < = 1;
@sum(year(j) : x(4,j))= 1; x(4,1)= 0; x(4,2)= 0; x(4,5)= 0;
@for(year(j): @sum(project(i):cost(i) * x(i,j)) < = yu(j));

End
```

LINGO 程序非零决策变量结果：

```
X(1, 3)      1.000000          0.000000
X(2, 4)      0.2250000         0.000000
X(2, 5)      0.2500000E- 01    0.000000
X(3, 1)      0.2000000         0.000000
X(3, 2)      0.4000000         0.000000
X(3, 3)      0.5333333E- 01    0.000000
X(3, 4)      0.3466667         0.000000
X(4, 3)      1.000000          0.000000
```

由上述 $x(i,j)$ 可知工程安排进度表，已达到五年内最大总收入 523.75 万元。

3. 广告宣传优化模型

(1) 问题介绍

某公司是一家小型公司，它最近要将一种新的旅游产品投放到北京地区的市场中，并且希望通过各种媒体对此产品进行宣传。因此，该公司与当地一家专门从事地区宣传的广告代理商进行了联系，并将此任务完全托付给对方，预算金额总额为 250 000 元。此广告代理商对北京市场有充分的了解，推荐了 6 种不同的媒体，建议进行为期两个月的宣传；对于每种媒体，广告代理商都了解在其上做广告的成本，以及该媒体能够影响到的人数。此外，广告代理商估计了每种媒体的影响指数，并了解到每种媒体的最大使用量。表 5-12 中列出了上述信息。

表 5-12　媒体广告详细信息(问题介绍)

编号	媒 体 类 型	影响人数/人	单位费用/元	最大使用量	影响指数
1	《精品购物指南》周报	12 000	1 500	4 个星期	3
2	《旅游》月刊	1 500	8 000	2 个月	7
3	《都市快报》周刊	2 000	12 000	8 个星期	8
4	北京城市广播	6 000	9 000	60 次广播	2
5	户外广告牌	3 000	24 000	4 个广告牌	6

续表

编号	媒 体 类 型	影响人数/人	单位费用/元	最大使用量	影响指数
6	北京电视台生活频道	9 000	51 000	8 次播放	9

问题 1：该公司希望此广告宣传计划至少能够影响到 100 000 人。应该如何选择媒体组合，才能使广告的总影响指数最高？

问题 2：此问题和实际广告投放有一定的差距。你认为主要的差距有哪些？模型应如何改进？

（2）建立模型

① 模型假设

- 所有宣传活动全部在两个月内进行。
- 不同类型的媒体可同时进行宣传。
- 媒体宣传在两个月的时间内完成对应的宣传次数。

② 变量符号说明表

- X_i：表示第 i 个媒体的刊登或播放次数（$i=1,2,\cdots,6$）。
- P_i：表示第 i 个媒体的影响总人数（$i=1,2,\cdots,6$）。
- M_i：表示第 i 个媒体的使用总费用（$i=1,2,\cdots,6$）。
- E_i：表示第 i 个媒体的影响总指数（$i=1,2,\cdots,6$）。
- Z：表示媒体组合的最高总指数。

③ 模型建立

a. 设 Z 为最高总影响指数，建立目标函数

$$\max Z = 3X_1 + 7X_2 + 8X_3 + 2X_4 + 6X_5 + 9X_6$$
$$(X_i \in \mathbf{N},\ i = 1,2,\cdots,6)$$

b. 模型中，$X_1 \sim X_6$ 为决策变量，取值均可为 0。为 0 时，代表不采用对应媒体的宣传投放。

c. 根据企业要求，列出约束条件。

刊登或播放次数限制为

$$X_1 \leqslant 4,\quad X_2 \leqslant 2,\quad X_3 \leqslant 8,\quad X_4 \leqslant 60,\quad X_5 \leqslant 4,\quad X_6 \leqslant 8$$

设 P_i 为广告影响总人数（$P_i \in \mathbf{N}^*$；$i=1,2,\cdots,6$），则广告使用量与影响人数的关系表达式为

$$P_1 = 12\ 000 \times X_1,\quad P_2 = 1\ 500 \times X_2,\quad P_3 = 2\ 000 \times X_3,$$
$$P_4 = 6\ 000 \times X_4,\quad P_5 = 3\ 000 \times X_5,\quad P_6 = 9\ 000 \times X_6$$

因为广告使用量（即使用时间或次数）决定消耗的总费用，所以，设 M_i（$i=1,2,\cdots,6$）为广告所需总费用，则广告使用量与费用的关系表达式为

$$M_1 = 1\ 500 \times X_1,\quad M_2 = 8\ 000 \times X_2,\quad M_3 = 12\ 000 \times X_3,$$
$$M_4 = 9\ 000 \times X_4,\quad M_5 = 24\ 000 \times X_5,\quad M_6 = 51\ 000 \times X_6$$

公司希望此广告宣传计划至少能够影响到 10 万人，则

$$P_1 + P_2 + P_3 + P_4 + P_5 + P_6 \geqslant 100\ 000$$

因为公司预算金额总额为 250 000 元，则

$$M_1 + M_2 + M_3 + M_4 + M_5 + M_6 \leqslant 250\ 000$$

d. 总模型

$$\max Z = 3X_1 + 7X_2 + 8X_3 + 2X_4 + 6X_5 + 9X_6$$
$$(X_i \in \mathbf{N},\ i = 1,2,\cdots,6)$$
$$P_1 + P_2 + P_3 + P_4 + P_5 + P_6 \geqslant 100\ 000$$
$$M_1 + M_2 + M_3 + M_4 + M_5 + M_6 \leqslant 250\ 000$$
$$X_1 \leqslant 4,\quad X_2 \leqslant 2,\quad X_3 \leqslant 8,\quad X_4 \leqslant 60,\quad X_5 \leqslant 4,\quad X_6 \leqslant 8$$

④ 模型求解

LINGO 程序如下所示。

```
model:
max=z;
z=3*x1+7*x2+8*x3+2*x4+6*x5+9*x6;
12000*x1+1500*x2+2000*x3+6000*x4+3000*x5+9000*x6>=100000;
1500*x1+8000*x2+12000*x3+9000*x4+24000*x5+51000*x6<=250000;
@bnd(0,x1,4);
@bnd(0,x2,2);
@bnd(0,x3,8);
@bnd(0,x4,60);
@bnd(0,x5,4);
@bnd(0,x6,8);
end
```

程序运行结果：

```
max Z=122
X1=4  X2=2  X3=8  X4=4  X5=4  X6=0
```

⑤ 结果分析

如表 5-13 所示，周报刊登 4 个星期，月刊刊登 2 个月，周刊刊登 8 个星期，电台广播 4 次，广告牌设置 4 个，不采用电视广告的方式来宣传。

表 5-13　广告投放影响表

媒体类型	影响人数/千人	总费用/千元	使用量	影响指数
周报	48	6	4 个星期	12
月刊	3	16	2 个月	14
周刊	16	96	8 个星期	64
电台广播	24	36	4 次广播	8
户外广告牌	12	96	4 个广告牌	24
电视台	0	0	0 次播放	0
合　计	103	250	22	122

最大总影响指数为 122，总影响人数最高达 103 000 人，消耗的费用恰好等于总预算 250 000 元。

采用的五种媒体中：周报的总影响人数最多，48 000 人；周刊与户外广告牌消耗费用最多，同为 96 000 元；周刊的总影响指数最高，为 64。周报价格最低，而影响人数最多；周刊的

价格虽然很高，但影响指数达到最大；户外广告牌的高费用消耗，却没有得到相应的影响人数和影响指数。

⑥ 进一步的思考

该模型的情况较真实问题简化了很多。比如，电台广告的有效性不大可能与广告的播放次数成正比，不同媒体上的广告之间也不太可能毫无关系。

4. 妈妈的理财计划——字斟句酌读懂题目内涵

(1) 问题介绍

妈妈希望儿子利用计算机帮助自己进行每年的家庭理财规划。她列出每个月的支出和收入：每月生活费 5 500 元，房租 6 300 元；电话费每 2 个月 1 350 元，燃气、电费及上网费每 6 个月 8 500 元，汽车每个月 3 400 元；纳税每 4 个月 1 000 元；收入每个月 19 000 元工资，加上 1 500 元的补贴。妈妈每个月至少要花 1 650 元用于各种娱乐，她希望能有更多的娱乐。那么，应该怎样安排每年的预算，才能使用于娱乐的预算最多？

(2) 建立模型

① 假设

a. 最初积蓄余额为 0。

b. 2 月期的支出只计入偶数月的预算支出。

c. 4 月期的支出只计入 4 月、8 月和 12 月的预算支出。

d. 6 月期的支出只计入 6 月和 12 月的预算支出。

为了家庭生活正常，必须平衡收支。每个月的支出不能超过当月的收入，但可以保存部分收入，以应对某个月开销较大的情况。

② 决策变量

$x_i(i=1,2,\cdots,12)$表示第 i 个月的娱乐花费。

注意：可以引进一组变量表示每个月的积蓄。这样，模型表示起来比较清楚。

③ 约束条件

a. 每个月至少要花 1 650 元用于各种娱乐。

b. 每个月的支出小于等于收入。

④ 决策目标

最大化每年可以用于娱乐的总金额。

⑤ 结论

一年中，该家庭用于娱乐的支出共计 35 500 元。表 5-14 列出了每个月的娱乐预算方案(有多个等价的方案)。

表 5-14 每月娱乐支出与积蓄值

月　份	1	2	3	4	5	6	7	8	9	10	11	12
娱乐支出/元	5 300	3 950	4 050	1 650	1 650	1 650	5 300	2 950	4 050	1 650	1 650	1 650
积蓄/元	0	0	1 250	2 550	6 200	0	0	0	1 250	3 550	7 200	0

5. 员工时序安排模型

(1) 问题分析

某公司客服部需要在网络和电话热线上接待顾客，解答疑问等。根据服务数据历史信息统计可知，一周中每天需要不同数量的雇员：周一到周四每天至少需要 5 人，周五需要 8 人，周六和周日需要 10 人。现为了工作安排方便，规定应聘者需连续工作 5 天，然后休息 2 天。试确定聘用方案，使得在满足需要的前提下，聘用总人数最少。

(2) 建立模型

① 决策变量

记周一到周日每天开始工作的人数分别为 $x_1, x_2, x_3, x_4, x_5, x_6, x_7$。

② 目标函数

目标函数即聘用总人数 Z

$$Z = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7$$

③ 约束条件

由每天需要的人数确定。

由于每人连续工作 5 天，所以周一的雇员应该是周四到周一开始上班的人，按照需要，至少 5 人。于是

$$x_1 + x_4 + x_5 + x_6 + x_7 \geqslant 5$$

类似地，有

$$x_1 + x_2 + x_5 + x_6 + x_7 \geqslant 5$$
$$x_1 + x_2 + x_3 + x_6 + x_7 \geqslant 5$$
$$x_1 + x_2 + x_3 + x_4 + x_7 \geqslant 5$$
$$x_1 + x_2 + x_3 + x_4 + x_5 \geqslant 8$$
$$x_2 + x_3 + x_4 + x_5 + x_6 \geqslant 10$$
$$x_3 + x_4 + x_5 + x_6 + x_7 \geqslant 10$$

显然，人数应该是非负整数，所以

$$x_i \geqslant 0 \quad (i = 1, 2, \cdots, 7)$$

问题归结为：在以上约束条件下，求解 $\min Z$ 的整数规划模型。由于目标函数和约束条件关于决策变量都是线性函数，所以这是一个整数线性规划模型。

④ LINGO 程序求解

```
model:
sets:
day/1..7/:x;
endsets
min=@sum(day : x);
!@for(day : @gin(x));
x1+x4+x5+x6+x7>=5;
x1+x2+x5+x6+x7>=5;
x1+x2+x3+x6+x7>=5;
x1+x2+x3+x4+x7>=5;
x1+x2+x3+x4+x5>=8;
x2+x3+x4+x5+x6>=10;
x3+x4+x5+x6+x7>=10;
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

$$x(i) \quad (i = 1,2,\cdots,7)$$

[illegible]

$$\min Z = [illegible]$$

[illegible]

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

[illegible]

$$[illegible]$$

[illegible]

[illegible]

[illegible]

$$[illegible]$$

[illegible]

a. [illegible]

$$[illegible]$$

b. [illegible]

$$[illegible]$$

$$x(i) \in N, \quad i = 1,2,\cdots,7$$

[illegible] LINGO [illegible]

```
[illegible]
```

```
!每天所需的最少员工数;
enddata
min=z;
z=@sum(day(i):x(i));
@for(day(i):z-x(@wrap(i+1,7))-x(@wrap(i+2,7))>=n(i));
!@for(day(i):@gin(x(i)));
end
```

6. 混合泳问题

(1) 问题介绍

某班准备从5名游泳队员中选择4人组成接力队,参加学校的4×100米混合泳接力比赛。5名队员4种泳姿的百米平均成绩如表5-15所示,问应该如何选拔队员组成接力队?

表5-15　5名队员4种泳姿百米平均成绩(1)

队员	甲	乙	丙	丁	戊
蝶泳	1′06″8	57″2	1′18″	1′10″	1′07″4
仰泳	1′15″6	1′06″	1′07″8	1′14″2	1′11″
蛙泳	1′27″	1′06″4	1′24″6	1′09″6	1′23″8
自由泳	58″6	53″	59″4	57″2	1′02″4

(2) 建立模型

① 问题分析

要求从5名队员中选出4人组成接力队,每人1种泳姿,且4人的泳姿各不相同,使接力队成绩最好。容易想到穷举法,组成接力队的方案有5!=120种。逐一计算并做比较,可找出最优方案。显然,这不是解决问题的最好方法。随着问题规模变大,穷举法的计算量是无法接受的。

可以用0—1变量表示一个队员是否入选接力队,从而建立问题的0—1规划模型。

记甲、乙、丙、丁、戊分别为队员 $i=1,2,3,4,5$;记蝶泳、仰泳、蛙泳、自由泳分别为泳姿 $j=1,2,3,4$;记队员 i 的第 j 种泳姿的百米成绩为 c_{ij}(s),则成绩如表5-16所示。

表5-16　5名队员4种泳姿百米平均成绩(2)　　单位:s

队员	甲	乙	丙	丁	戊
蝶泳	66.8	57.2	78	70	67.4
仰泳	75.6	66	67.8	74.2	71
蛙泳	87	66.4	84.6	69.6	83.8
自由泳	58.6	53	59.4	57.2	62.4

② 决策变量

引入0—1变量 x_{ij}。若选择队员 i 参加泳姿 j 的比赛,记 $x_{ij}=1$,否则记 $x_{ij}=0$。这就是问题的决策变量,共20个。

③ 目标函数

当队员 i 入选泳姿 j 的比赛时,$c_{ij}x_{ij}$ 表示他的成绩,否则 $c_{ij}x_{ij}=0$。于是,接力队的成绩

表示为

$$f = \sum_{j=1}^{4} \sum_{i=1}^{5} c_{ij} x_{ij}$$

④ 约束条件

根据组成接力队的要求，x_{ij}应该满足下面两个约束条件。

a. 每人最多只能选择 4 种泳姿之一，即对于队员 $i=1,2,3,4$，应该有

$$\sum_{j=1}^{4} x_{ij} \leqslant 1$$

b. 每种泳姿有且只能有 1 人入选，即对于队员 $j=1,2,3,4,5$，应该有

$$\sum_{i=1}^{5} x_{ij} = 1$$

综上所述，这个问题的优化模型可以写作：

$$\min \sum_{j=1}^{4} \sum_{i=1}^{5} c_{ij} x_{ij}$$

$$\text{s. t.} \begin{cases} \sum_{j=1}^{4} x_{ij} \leqslant 1 \quad (i = 1,2,3,4,5) \\ \sum_{i=1}^{5} x_{ij} = 1 \quad (j = 1,2,3,4) \\ x_{ij} = \{0,1\} \end{cases}$$

这是一个线性 0—1 规划模型。它是一个特殊的线性整数规划。

学习信息库 3——LINGO 错误编号及含义对照表

在运用 LINGO 程序求解问题时，系统首先会对程序进行编译。系统在编译或执行其他命令时，会因程序中的错误或运行错误，弹出出错报告窗口，显示错误代码，并简要指出错误的原因。这些错误报告信息能够提示用户发现程序中的错误，以便尽快修改。表 5-17 给出出错信息的简要说明，仅供参考。

表 5-17　错误代码及含义

错误代码	含　义
0	LINGO 模型生成器的内存已经用尽(可用 LINGO→Options 命令对 General Solver 选项卡中的 Generator Memory Limit 选项进行内存大小的修改)
1	模型中的行数太多(对于有实际意义的模型，这个错误很少出现)
2	模型中的字符数太多(对于有实际意义的模型，这个错误很少出现)

续表

错误代码	含　义
3	模型中某行的字符数太多(每行不应该超过 200 个字符,否则应换行)
4	指定的行号超出了模型中实际具有的最大行号(这个错误通常在 LOOK 命令中指定了非法的行号时出现)
5	当前内存中没有模型
6	脚本文件中,TAKE 命令的嵌套重数太多(LINGO 中,限定 TAKE 命令最多嵌套 10 次)
7	无法打开指定的文件(通常是指定的文件名拼写错误)
8	脚本文件中的错误太多,因此直接返回到命令模式(不再继续处理这个脚本文件)
9	(该错误编号目前没有使用)
10	(该错误编号目前没有使用)
11	模型中的语句出现了语法错误(不符合 LINGO 语法)
12	模型中的括号不匹配
13	在电子表格文件中找不到指定的单元范围名称
14	运算所需的,临时堆栈空间不够(这通常意味着模型中的表达式太长了)
15	找不到关系运算符(通常是丢了“＜”“＝”或“＞”)
16	输入、输出时,不同对象的大小不同(使用集合循环方式输入、输出时,集合大小应相同)
17	集合元素的索引的内存堆栈空间不够
18	集合的内存堆栈空间不够
19	索引函数@INDEX 使用不当
20	集合名使用不当
21	属性名使用不当
22	不等式或等式关系太多(例如,约束 $2<x<4$ 是不允许出现在同一条语句中的)
23	参数个数不符
24	集合名不合法
25	函数@WKX()的参数非法(注:在 LINGO 9.0 中已经没有函数@WKX())
26	集合的索引变量的个数不符
27	在电子表格文件中指定的单元范围不连续
28	行名不合法
29	数据段或初始段的数据个数不符
30	链接到 Excel 时出现错误
31	使用@TEXT 函数时,参数不合法
32	使用了空的集合成员名
33	使用@OLET 函数时,参数不合法
34	用电子表格文件中指定的多个单元范围生成派生集合时,单元范围的大小应该一致
35	输出时用到了不可识别的变量名
36	基本集合的元素名不合法
37	集合名已经被使用过
38	ODBC 服务返回了错误信息
39	派生集合的分量元素(下标)不在原来的父集合中

续表

错误代码	含　义
40	派生集合的索引元素的个数不符
41	定义派生集合时所使用的基本集合的个数太多(一般不会出现这个错误)
42	集合过滤条件的表达式中出现了取值不固定的变量
43	集合过滤条件的表达式运算出错
44	过滤条件的表达式没有结束(即没有“:”标志)
45	@ODBC 函数的参数列表错误
46	文件名不合法
47	打开的文件太多
48	不能打开文件
49	读文件时发生错误
50	@FOR 函数使用不合法
51	编译时,LINGO 模型生成器的内存不足
52	@IN 函数使用不当
53	在电子表格文件中找不到指定的单元范围名称(与出错代码“13”含义类似)
54	读取电子表格文件时出现错误
55	@TEXT 函数不能打开文件
56	@TEXT 函数读文件时发生错误
57	@TEXT 函数读文件时出现了非法输入数据
58	@TEXT 函数读文件时,发现输入数据比实际所需要的少
59	@TEXT 函数读文件时,发现输入数据比实际所需要的多
60	用@TEXT 函数输入数据时,没有指定文件名
61	行命令拼写错误
62	LINGO 生成模型时,工作内存不足
63	模型的定义不正确
64	@FOR 函数嵌套太多
65	@WARN 函数使用不当
66	警告:固定变量取值不唯一(例如,任意正数都是约束@SIGN(X)=1 的解)
67	模型中,非零系数过多,导致内存耗尽
68	对字符串进行非法的算术运算
69	约束中的运算符非法
70	属性的下标越界
71	变量定界函数(@GIN、@BIN、@FREE、@BND)使用错误
72	不能从固定约束(只含有固定变量的约束)中求出固定变量的值(相当于方程无解,或者 LINGO 的算法解不出来,如迭代求解算法不收敛)
73	在 LINGO 生成模型(对模型进行结构分析)时,用户中断了模型生成过程
74	变量越界,超出了 1032
75	对变量的定界相互冲突(例如,一个模型中同时指定@BND(-6,X,6)和@BND(-5,X,5)是允许的,但同时指定@BND(-6,X,6)和@BND(7,X,9)是冲突的)
76	LINGO 生成模型时出现错误,不能将模型转交给优化求解程序

续表

错误代码	含　义
77	无定义的算术运算(例如,除数为0)
78	(该错误编号目前没有使用)
79	(该错误编号目前没有使用)
80	生成LINGO模型时,系统内存已经用尽
81	找不到可行解
82	最优值无界
83	(该错误编号目前没有使用)
84	模型中非零系数过多
85	表达式过于复杂,导致堆栈溢出
86	算术运算错误(如1/0或@LOG(－1)等)
87	@IN函数使用不当(似乎与错误代码"52"相同)
88	当前内存中没有存放任何解
89	LINGO运行时出现了意想不到的错误(请与LINGO公司联系,解决问题)
90	在LINGO生成模型时,用户中断了模型生成过程
91	当在数据段有"变量＝?"语句时,LINGO运行时将要求用户输入这个变量的值。如果该值输入错误,将显示这个错误代码
92	警告:当前解可能不是可行的/最优的
93	命令行中的转换修饰词错误
94	(该错误编号目前没有使用)
95	命令行中的转换修饰词错误
96	(该错误编号目前没有使用)
97	用TAKE命令输入模型时,出现了不可识别的语法
98	用TAKE命令输入模型时,出现了语法错误
99	语法错误,缺少变量
100	语法错误,缺少常量
101	(该错误编号目前没有使用)
102	指定的输出变量名不存在
103	(该错误编号目前没有使用)
104	模型还没有被求解,或者模型是空的
105	(该错误编号目前没有使用)
106	行宽的最小值和最大值分别为68和200
107	函数@POINTER指定的索引值无效
108	模型的规模超出了当前LINGO版本的限制
109	达到了迭代上限,所以LINGO停止继续求解模型(迭代上限可以通过LINGO→Options命令对General Solver选项卡中的Iteration选项进行修改)
110	HIDE(隐藏)命令指定的密码超出了8个字符的限制
111	模型是隐藏的,所以当前命令不能使用
112	恢复隐藏模型时输入的密码错误
113	因为一行内容太长,导致LOOK或SAVE命令失败

续表

错误代码	含　义
114	HIDE(隐藏)命令指定的两次密码不一致,命令失败
115	参数列表过长
116	文件名(包括路径名)太长
117	无效的命令
118	命令不明确(例如,可能输入的是命令的缩写名,该缩写可有多个命令与之对应)
119	命令脚本文件中的错误太多,LINGO 放弃对它继续处理
120	LINGO 无法将配置文件(LINGO. CNF)写入启动目录或工作目录(可能是权限问题)
121	整数规划没有敏感性分析
122	敏感性分析选项没有激活,敏感性分析不能进行(可通过 LINGO→Options 命令对 General Solver 选项卡中的 Dual Computation 选项进行修改)
123	调试(Debug)命令只对线性模型,且模型不可行或无界时才能使用
124	对一个空集合的属性进行初始化
125	集合中没有元素
126	使用 ODBC 连接输出时,发现指定的输出变量名不存在
127	使用 ODBC 连接输出时,同时输出的变量的维数必须相同
128	使用 SET 命令时,指定的参数索引无效
129	使用 SET 命令时,指定的参数的取值无效
130	使用 SET 命令时,指定的参数名无效
131	FREEZE 命令无法保存配置文件 LINGO. CNF(可能是权限问题)
132	LINGO 读配置文件(LINGO. CNF)时发生错误
133	LINGO 无法通过 OLE 连接电子表格文件(例如,当其他人正在编辑这个文件时)
134	输出时出现错误,不能完成所有输出操作
135	求解时间超出了限制(可通过 LINGO→Options 命令对 General Solver 选项卡中的 Time 选项进行修改)
136	使用@TEXT 函数输出时出现错误操作
137	(该错误编号目前没有使用)
138	DIVERT(输出重新定向)命令的嵌套次数太多(最多不能超过 10 次嵌套)
139	DIVERT(输出重新定向)命令不能打开指定文件
140	只求原始最优解时,无法给出敏感性分析信息(可通过 LINGO→Options 命令对 General Solver 选项卡中的 Dual Computation 选项进行修改)
141	对某行约束的敏感性分析无法进行,因为这一行已经是固定约束(即该约束中的所有变量都已经在直接求解程序预处理时被固定下来了)
142	出现了意想不到的错误(请与 LINDO 公司联系解决)
143	使用接口函数输出时,同时输出的对象的维数必须相同
144	@POINTER 函数的参数列表无效
145	@POINTER 函数出错:2——输出变量无效;3——内存耗尽;4——只求原始最优解时,无法给出敏感性分析信息;5——对固定行无法给出敏感性分析信息;6——意想不到的错误
146	基本集合的元素名与模型中的变量名重名(当前版本的 LINGO 中,这本来是允许的,但如果通过 LINGO→Options 命令在 General Solver 选项卡选择 Check for duplicates names in data and model,会检查重名。这主要是为了与以前的 LINGO 版本兼容)

续表

错误代码	含　义
147	@WARN 函数中的条件表达式中只能包含固定变量
148	@OLE 函数在当前操作系统下不能使用(只在 Windows 操作系统下可以使用)
149	(该错误编号目前没有使用)
150	@ODBC 函数在当前操作系统下不能使用(只在 Windows 操作系统下可以使用)
151	@POINTER 函数在当前系统下不能使用(只在 Windows 操作系统下可以使用)
152	输入的命令在当前操作系统下不能使用
153	集合的初始化(定义元素)不能在初始段中进行,只能在集合段或数据段进行
154	集合名只能被定义一次
155	在数据段对集合进行初始化(定义元素)时,必须显式地列出所有元素,不能省略元素
156	在数据段对集合和(或)变量初始化时,给出的参数个数不符
157	@INDEX 函数引用的集合名不存在
158	当前函数需要集合的成员名作为参数
159	派生集合中的一个成员(分量)不是对应的父集合的成员
160	数据段中的一个语句不能对两个(或更多)集合初始化(定义元素)
161	(该错误编号目前没有使用)
162	电子表格文件指定的单元范围内存在不同类型的数据(既有字符,又有数值),LINGO 无法通过这些单元同时输入(或输出)不同类型的数据
163	在初始段对变量初始化时,给出的参数个数不符
164	模型中输入的符号名不符合 LINGO 的命名规则
165	当前的输出函数不能按集合输出
166	不同长度的输出对象无法同时输出到表格型的文件(如数据库和文本文件)
167	在通过 Excel 输入、输出时,一次指定了多个单元范围
168	@DUAL、@RANGEU、@RANGED 函数不能对文本数据(如集合的成员名)使用,只能对变量和约束行使用
169	运行模型时才输入集合成员是不允许的
170	LINGO 系统的密码输入错误,请重新输入
171	LINGO 系统的密码输入错误,系统将以演示版方式运行
172	LINGO 的内部求解程序发生了意想不到的错误(请与 LINDO 公司联系解决问题)
173	内部求解程序发生了数值计算方面的错误
174	LINGO 预处理阶段(preprocessing)内存不足
175	系统的虚拟内存不足
176	LINGO 后处理阶段(postprocessing)内存不足
177	为集合分配内存时出错(如内存不足等)
178	为集合分配内存时,堆栈溢出
179	将 MPS 格式的模型文件转化成 LINGO 模型文件时出现错误(如变量名冲突等)
180	将 MPS 格式的模型文件转化成 LINGO 模型文件时,不能分配内存(通常是内存不足)
181	将 MPS 格式的模型文件转化成 LINGO 模型文件时,不能生成模型(通常是内存不足)
182	将 MPS 格式的模型文件转化成 LINGO 模型文件时出现错误(会给出出错的行号)
183	LINGO 目前不支持 MPS 格式的二次规划模型文件

续表

错误代码	含　　义
184	敏感性分析选项没有激活，敏感性分析不能进行（可通过 LINGO→Options 命令对 General Solver 选项卡中的 Dual Computation 选项进行修改）
185	没有使用内点法的权限（LINGO 中的内点法是选件，需要额外购买）
186	不能用@QRAND 函数对集合初始化（定义元素）
187	用@QRAND 函数对属性初始化时，一次只能对一个属性进行处理
188	用@QRAND 函数对属性初始化时，只能对稠密集合对应的属性进行处理
189	随机函数中指定的种子（SEED）无效
190	用隐式方法定义集合时，定义方式不正确
191	LINDOAPI 返回了错误（请与 LINDO 公司联系解决问题）
192	LINGO 不再支持@WKX 函数，请改用@OLE 函数
193	内存中没有当前模型的解（模型可能还没有求解，或者求解错误）
194	无法生成 LINGO 的内部环境变量（通常是因为内存不足）
195	写文件时出现错误（如磁盘空间不足）
196	无法为当前模型计算对偶解（这个错误非同寻常，欢迎将该模型提供给 LINDO 公司进一步分析）
197	调试程序目前不能处理整数规划模型
198	当前二次规划模型不是凸的，不能使用内点法，请通过 LINGO→Options 命令取消对二次规划的判别
199	求解二次规划需要使用内点法，但使用的 LINGO 版本没有这个权限（请通过 LINGO→Options 命令取消对二次规划的判别）
200	无法为当前模型计算对偶解，请通过 LINGO→Options 命令取消对对偶计算的要求
201	模型是局部不可行的
202	全局优化时，模型中非线性变量的个数超出了全局优化求解程序的上限
203	无权使用全局优化求解程序
204	无权使用多初始点求解程序
205	模型中的数据不平衡（数量级差异太大）
206	“线性化”和“全局优化”两个选项不能同时存在
207	缺少左括号
208	@WRITEFOR 函数只能在数据段出现
209	@WRITEFOR 函数中不允许出现关系运算符
210	@WRITEFOR 函数使用不当
211	输出操作中出现了算术运算错误
212	集合的下标越界
213	当前操作参数不应该是文本，但模型中指定的是文本
214	多次对同一个变量初始化
215	@DUAL、@RANGEU、@RANGED 函数不能在此使用（参阅错误代码“168”）
216	这个函数应该需要输入文本作为参数
217	这个函数应该需要输入数值作为参数

续表

错误代码	含　义
218	这个函数应该需要输入行名或变量名作为参数
219	无法找到指定的行
220	没有定义的文本操作
221	@WRITE 或@WRITEFOR 函数的参数溢出
222	需要指定行名或变量名
223	向 Excel 文件写数据时，动态接收单元超出了限制
224	向 Excel 文件写数据时，需要写的数据的个数多于指定的接收单元的个数
225	计算段(CALC)的表达式不正确
226	不存在默认的电子表格文件，请为@OLE 函数指定一个电子表格文件
227	为 APISET 命令指定的参数索引不正确
228	通过 Excel 输入、输出数据时，如果 LINGO 中的多个对象对应于 Excel 中的名，则列数应该一致
229	为 APISET 命令指定的参数类型不正确
230	为 APISET 命令指定的参数值不正确
231	APISET 命令无法完成
232~	(该错误编号目前没有使用)
1000	(错误编号为 1000 以上的信息，只对 Windows 系统有效)
1001	LINGO 找不到与指定括号匹配的括号
1002	当前内存中没有模型，不能求解
1003	LINGO 现在正忙，不能马上响应请求
1004	LINGO 不能写 LOG(日志)文件，也许磁盘已满
1005	LINGO 不能打开指定的 LOG(日志)文件
1006	不能打开文件
1007	没有足够内存完成命令
1008	不能打开新窗口(可能内存不够)
1009	没有足够内存空间生成解答报告
1010	不能打开 Excel 文件的链接(通常是由于系统资源不足)
1011	LINGO 不能完成对图形的请求
1012	LINGO 与 ODBC 连接时出现错误
1013	通过 OBDC 传递数据时，不能完成初始化
1014	向 Excel 文件传递数据时，指定的参数不够
1015	不能保存文件
1016	Windows 环境下不支持 Edit 命令，请使用 File→Open 菜单命令
9999	由于出现严重错误，优化求解程序运行失败(最可能的原因是数学函数出错。如函数@LOG($X-1$)，当 $X\leqslant 1$ 时，会出现这类错误)

学习信息库 4——优秀论文

NBA 赛程的分析与评价

（2008 年全国大学生数学建模竞赛 D 题）

摘要：本文分析赛程对球队的主要影响因素，即背靠背、连续客场作战、比赛时间间隔、连续遭遇强队相遇四种因素。采用层次分析法，对 NBA 赛程进行分析，得出赛程对球队的利弊的数量指标。根据数量指标，分析出这四种因素对于火箭队的影响，并对模型进行检验。用 0—1 规划方法建立同区不同区域球队比赛中，选取赛 3 场的球队安排的优化模型，并用 LINGO 软件求解。

本文的特色在于应用改进的层次分析法对赛程进行评价，同时应用 MATLAB 强大的矩阵计算功能，简化了分析求解过程，并且巧妙地将赛程安排问题转化为具有线性条件的 0—1 规划问题。

关键词：层次分析法 MATLAB　LINGO　0—1 规划

一、问题重述

对于 NBA 这样一个规模庞大的赛事，编制一个完整的、对各队尽可能公平的赛程是一件很复杂的事情。赛程安排的好坏，直接影响球队水平的发挥。因此，对于编排赛程的人来说，要考虑有哪些影响赛程的因素，并且建立数学模型，评价赛程利弊的数量指标。根据建立的数学模型，分析赛程对火箭队的利弊，并算出 30 支球队中最不利和最有利的球队。另外，找出选取赛 3 场的球队的方法，并对这种方法给予评价，找出你认为更好的方法。

二、模型假设

(1) 比赛不因外界因素的干扰而中断，所有比赛按赛程进行。

(2) 忽略参赛队每天参加比赛的具体时间对比赛公平性的影响。

三、日期及队名的处理

1. 对比赛日期的处理

根据相关数据信息，对比赛日期进行编号处理。将 2008 年 10 月 29 日，即比赛的第一天，记为“1”；之后每天的编号为前一天的编号加 1。例如，2008 年 10 月 30 日记为“2”，10 月 31 日记为“3”，以此类推。

2. 对球队名称的处理

把 30 支球队的名称分别从 1 到 30 编号，如表 1 所示。

表 1　球队及编号

东部联盟(Eastern conference)		西部联盟(Western conference)	
大西洋赛区		太平洋赛区	
波士顿凯尔特人(Boston)	1	洛杉矶湖人(L. A. lakers)	16

续表

东部联盟(Eastern conference)		西部联盟(Western conference)	
大西洋赛区		太平洋赛区	
新泽西网(New Jersey)	2	萨克拉门托国王(Sacramento)	17
纽约尼克斯(New York)	3	菲尼克斯太阳(Phoenix)	18
费城76人(Philadelphia)	4	金州勇士(Golden State)	19
多伦多猛龙(Toronto)	5	洛杉矶快船(L. A. Clippers)	20
中部赛区		西北赛区	
底特律活塞(Detroit)	6	明尼苏达森林狼(Minnesota)	21
印第安纳步行者(Indiana)	7	犹他爵士(Utah)	22
密尔沃基雄鹿(Milwaukee)	8	丹佛掘金(Denver)	23
芝加哥公牛(Chicago)	9	波特兰开拓者(Portland)	24
克里夫兰骑士(Cleveland)	10	西雅图超音速(Oklahoma City)	25
东南赛区		西南赛区	
迈阿密热火(Miami)	11	新奥尔良黄蜂(New Orleans)	26
奥兰多魔术(Orlando)	12	达拉斯小牛(Dallas)	27
华盛顿奇才(Washington)	13	圣安东尼奥马刺(San Antonio)	28
亚特兰大老鹰(Atlanta)	14	休斯敦火箭(Houston)	29
夏洛特山猫(Charlotte)	15	孟菲斯灰熊(Memphis)	30

四、模型的建立和求解

4.1 赛程对球队影响的因素

将当年NBA比赛开始的日期记为第1天,后面的日期以此类推,记编号为 n 的球队第 m 场($n=1,2,\cdots,30;m=1,2,\cdots,82$)比赛的天数序号为 $t_{n,m}$,比赛对手的编号为 $O_{n,m}$,主客场的信息为 $H_{n,m}$,且满足

$$H_{n,m}=\begin{cases}1, & \text{当球队 } n \text{ 的第 } m \text{ 场比赛为主场}\\ 0, & \text{当球队 } n \text{ 的第 } m \text{ 场比赛为客场}\end{cases}$$

则将影响每支球队比赛公平性的因素表述为如下几个方面。

1. 主客场因素

2008—2009 NBA赛季中,已知每个球队都要打82场比赛,所以仅考虑主客场数。一般而言,客场比赛的困难更大一些,为使每支球队有公平的竞争环境,应当满足每支球队客场作战41场。客场比赛场次越多,说明赛程对这支球队越不利。用球队 n 的客场作战场次来衡量主客场因素对它的影响,将主客场因素 $HA(n)$ 记为

$$HA(n)=|\,m\mid\{H_{n,m}=0,\forall\, m\}\,|$$

其中,$|m|$ 表示一个集合中元素的数目,下同。

通过统计数据得出的结论:2008—2009 NBA赛季中,每支球队在客场作战数是绝对相等的,各为41场。也就是说,单从主客场因素分析,说明赛程的安排对每支球队是公平的。因此,对此因素不作考虑。

2. 背靠背因素

“背靠背”的英文为 black-to-black,在职业联赛术语中,特指连续作战。“背靠背”的称

谓自由度极强，在不同场合也有不同的针对性。狭义的“背靠背”特指连续作战两个晚上，在不同的客场迎战不同对手；广义的“背靠背”可以理解为连续两个晚上在不同场地迎战不同的对手。背靠背，对球队来说，是“遍地荆棘”的代名词；对球员来说，无疑面临险恶的连番挑战。NBA 六百多名球员，喜怒哀乐各有不同，但可以肯定的是，没有任何一位球员喜欢打“背靠背”比赛，更别说连续的客场“背靠背”，这不是考验胜负，简直是在考验运气。所以，“背靠背”这个因素对赛程的影响是非常大的。

这里使用广义的“背靠背”概念，用整个赛程中球队遇到“背靠背”情况的次数，来衡量该因素对球队 n 的影响 $BB(n)$，即

$$BB(n)=|\{m \mid t_{n,m+1}-t_{n,m}=1,\ m=1,2,\cdots,81\}|$$

对于 2008—2009 赛季的 NBA 赛场，$BB(n)$ 的具体数值如表 2 所示。

表 2　NBA 赛场的 $BB(n)$ 数值

n	1	2	3	4	5	6
$BB(n)$	17	22	18	21	17	16
n	7	8	9	10	11	12
$BB(n)$	21	22	22	19	19	16
n	13	14	15	16	17	18
$BB(n)$	18	22	21	19	22	19
n	19	20	21	22	23	24
$BB(n)$	15	21	22	21	21	16
n	25	26	27	28	29	30
$BB(n)$	18	19	16	18	20	22

3. 连续客场作战因素

如果说“背靠背”是一个球队发挥好坏的重要指标之一，那么连续客场作战也必将成为关系球队胜率的另一杀手。客场连续作战不考虑比赛天数是否相连，但连续的客场作战使球队往返奔波于其他球队的主场之间，对体力和精神意志都是巨大的考验。这里用连续两场客场作战在赛程安排中的次数来衡量连续客场作战对球队 n 的影响 $AA(n)$，即

$$AA(n)=|\{m \mid H_{n,m}=0\ \&H_{n,m+1}=0,\ m=1,2,\cdots,81\}|$$

对于 2008—2009 赛季的 NBA 赛场，$AA(n)$ 的具体数值如表 3 所示。

表 3　NBA 赛场的 $AA(n)$ 数值

n	1	2	3	4	5	6
$AA(n)$	19	15	20	18	19	17
n	7	8	9	10	11	12
$AA(n)$	15	19	20	16	20	18
n	13	14	15	16	17	18
$AA(n)$	17	22	19	23	22	21
n	19	20	21	22	23	24
$AA(n)$	22	22	18	20	20	24
n	25	26	27	28	29	30
$AA(n)$	19	19	20	19	20	18

4. 比赛时间间隔因素

纵观往年赛季的赛程安排日期表，都会令一些球队不满意。有些球队被安排的日期前松后紧，或前紧后松，这些因素都会影响球员发挥，使球员的体力消耗不均匀，影响比赛成绩。记录球队 n 每相邻两场比赛的间隔时间为 $t_{n,m}^{(int)}=t_{n,m+1}-t_{n,m}(m=1,2,\cdots,81)$，用间隔时间的方差来表示比赛时间间隔因素对球队 n 的影响 $IN(n)$，即

$$IN(n)=\sigma_m(t_{n,m}^{(int)})$$

其中，σ_m 表示以 m 为变量求取方差。

对于 2008—2009 赛季的 NBA 赛场，$IN(n)$ 的具体数值如表 4 所示。

表 4　NBA 赛场的 $IN(n)$ 数值

n	1	2	3	4	5	6
$IN(n)$	0.924 6	0.971 8	0.891 3	1.022	0.918 9	0.802 8
n	7	8	9	10	11	12
$IN(n)$	0.891 3	0.951 3	0.911	0.951 3	0.877 2	0.833 3
n	13	14	15	16	17	18
$IN(n)$	0.891 3	0.971 8	0.899 2	0.918 9	1.058	0.932 4
n	19	20	21	22	23	24
$IN(n)$	0.818 2	0.997 2	0.945 8	0.885 2	1.022	0.839 6
n	25	26	27	28	29	30
$IN(n)$	0.848 2	0.932 4	0.841 8	0.891 3	0.891 3	0.945 8

5. 连续遭遇强队因素

连续遭遇强队是限制球队发挥的另一个决定因素。由于在面对强队比赛时，队员的体力消耗很大，势必影响下一场比赛的状态。如果连续面对强队比赛，对球队是非常严峻的考验。首先，定义上赛季常规赛胜率 $W(n)$ 大于某一阈值 W_{th} 的球队为强队，用在整个赛程中连续两次遭遇强队的次数来衡量连续遭遇强队对球队 n 的影响 $S(n)$，即

$$S(n)=|\{m\mid W(O_{n,m})>W_{th}\&W(O_{n,m+1})>W_{th},\ m=1,2,\cdots,81\}|$$

对于 2008—2009 赛季的 NBA 赛场，$S(n)$ 的具体数值如表 5 所示。

表 5　NBA 赛场的 $S(n)$ 数值

n	1	2	3	4	5	6
$S(n)$	11	19	16	15	12	15
n	7	8	9	10	11	12
$S(n)$	20	19	18	11	12	11
n	13	14	15	16	17	18
$S(n)$	10	16	14	18	24	17
n	19	20	21	22	23	24
$S(n)$	20	24	20	19	16	20
n	25	26	27	28	29	30
$S(n)$	20	18	16	21	18	23

4.2　数据属性值规范化

属性值的规范化用于解决属性之间不可同一度量的问题，及解决不同属性之间数值相

差悬殊的问题。本文中的四个因素数值相差悬殊，故采用下述方法规范化，使得数值变换到[0,1]之间。

有 n 个评价对象，记 i 是第 i 个评价对象属性数值。设 X_i 是 Y_i 的规范化属性值，故有

$$X_i = \frac{Y_i - \min_i\{Y_i\}}{\max_i\{Y_i\} - \min_i\{Y_i\}}$$

4.3 评价赛程利弊的数量指标

层次分析法的一般思路：把复杂的问题分解成因素；然后把这些因素按支配关系分组形成有序的递阶层次结构，并衡量各方面的影响；最后综合人的判断，决定决策诸因素相对重要性的先后、优劣次序。这就是层次分析法的基本思路。

(1) 分析系统中各因素间的关系，构造如图 1 所示的层次结构。

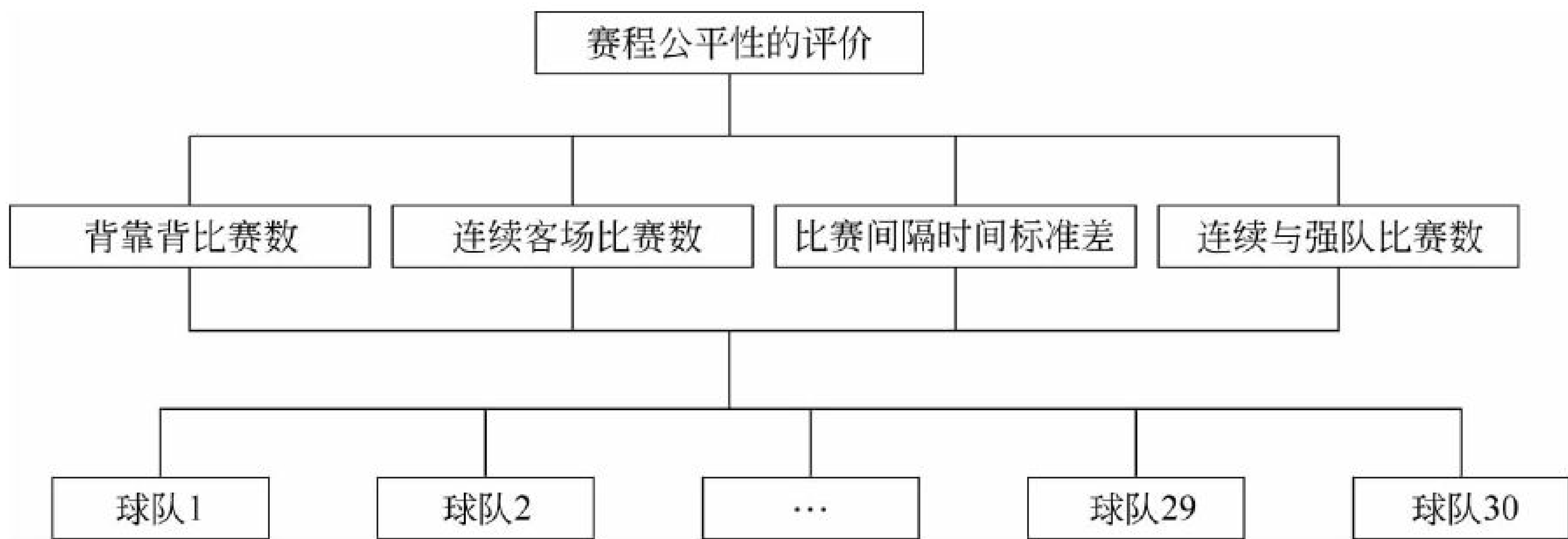


图 1　层次结构的设定

(2) 对同一层次的各元素关于上一层次中某一准则的重要性进行两两比较，构造两两比较的判断矩阵(或成对比矩阵)，做出权重分析。

对于本问题，通过与资深体育评论家以及查阅大量体育赛事的评论，如汪默(现供职于美国 NBA 总部)为 NBA 中国官方网站撰写评论及专栏文章，得出一个关于主客场、背靠背、连续客场作战、时间间隔、强强相遇四种因素的判断矩阵，如下所示。

$$\boldsymbol{A} = \begin{pmatrix} 1 & 3 & 1 & 1/3 \\ 1/3 & 1 & 1/5 & 1/3 \\ 1 & 5 & 1 & 3 \\ 3 & 3 & 1/3 & 1 \end{pmatrix}$$

(3) 计算一致性指标，检验一致性。

(4) 计算各层元素对于系统目标的总排序权重(即合成权重)及模型的决策。

设影响因素背靠背比赛数、连续客场比赛数、比赛间隔时间标准差、连续与强队比赛数分别为 p、q、r、s，利用如下方法构造各球队四个因素的判断矩阵 $\boldsymbol{B}$、$\boldsymbol{C}$、$\boldsymbol{D}$、$\boldsymbol{E}$。

$$b_{ij} = \frac{p_i}{p_j} = \frac{p_i}{p_k} \cdot \frac{p_k}{p_j} = b_{ik} \cdot b_{ki} \quad (i,j,k = 1,2,\cdots,30)$$

$$\boldsymbol{B} = (b_{ij})_{30\times30} = \left(\frac{p_i}{p_j}\right)_{30\times30}, \quad \boldsymbol{C} = \left(\frac{q_i}{q_j}\right)_{30\times30}, \quad \boldsymbol{D} = \left(\frac{r_i}{r_j}\right)_{30\times30}, \quad \boldsymbol{E} = \left(\frac{S_i}{S_j}\right)_{30\times30}$$

即矩阵 $\boldsymbol{B}$ 是一致性判断矩阵，$\boldsymbol{C}$、$\boldsymbol{D}$、$\boldsymbol{E}$ 同理。

一致性判断矩阵有如下性质：一致的正互反阵的秩等于1，主持正根为 n，若特征向量为 $W=(w_1,\cdots,w_n)'$，则有 $a_{ij}=w_i/w_j$。一致性判断矩阵的特征向量就是因素的排序，即 W 就

是各对象之间的一个排序。

根据以上算法，利用 MATLAB 软件，做出数量指标总排序如表 6 所示。

表 6 数量指标总排序

准则		Y_1	Y_2	Y_3	Y_4	数量指标
准则层权值		0.382 5	0.130 9	0.752 8	0.519 5	
球队赛程公平性影响因素的标准化值	1	0.833 3	0.555 6	0.522 6	0.928 6	11.192
	2	0.833 3	1	0.337 7	0.357 1	14.86
	3	0.833 3	0.444 4	0.653 1	0.571 4	13.895 3
	4	1	0.666 7	0.141 3	0.642 9	12.829 9
	5	1	0.555 6	0.544 9	0.857 1	11.324 8
	6	1	0.777 8	1	0.642 9	12.534
	7	1	1	0.653 1	0.285 7	14.936 4
	8	0.833 3	0.555 6	0.418 2	0.357 1	15.368
	9	1	0.444 4	0.576	0.428 6	14.566 6
	10	0.833 3	0.888 9	0.418 2	0.928 6	10.819 5
	11	0.666 7	0.444 4	0.708 5	0.857 1	12.189 1
	12	0.666 7	0.666 7	0.880 3	0.928 6	11.374 9
	13	0.833 3	0.777 8	0.653 1	1	10.385 7
	14	0.833 3	0.222 2	0.337 7	0.571 4	14.217 7
	15	0.833 3	0.555 6	0.622 1	0.714 3	12.731 4
	16	0.166 7	0.111 1	0.544 9	0.428 6	16.877 6
	17	0	0.222 2	0	0	20.350 9
	18	0	0.333 3	0.492	0.5	16.489
	19	0.166 7	0.222 2	0.939 6	0.285 7	17.709 9
	20	0	0.222 2	0.238 3	0	20.305 1
	21	0.5	0.666 7	0.439 9	0.285 7	16.517 4
	22	0.666 7	0.444 4	0.677	0.357 1	15.831 7
	23	0.833 3	0.444 4	0.141 3	0.571 4	13.993 7
	24	0.5	0	0.855 7	0.285 7	17.222 8
	25	0.5	0.555 6	0.822	0.285 7	16.574 9
	26	0.333 3	0.555 6	0.492	0.428 6	15.981 8
	27	1	0.444 4	0.847 1	0.571 4	13.475 5
	28	0.666 7	0.555 6	0.653 1	0.214 3	16.744 3
	29	0.333 3	0.444 4	0.653 1	0.428 6	16.081 7
	30	0.833 3	0.666 7	0.439 9	0.071 4	17.311

从上述图表可以得出，编号为 17 的数量指标最高，表示整个赛程的编排对它最公平，是最优的；编号为 13 的数量指标最低，表示整个赛程的编排对它是最不公平、最差的。

4.4 新赛季选取赛 3 场的球队的方法

经过网上查询资料，我们发现，赛 3 场的球队都是随机抽取的，即比赛双方是在一定条

件下，由计算机随机选择。这样，公平性大大降低，因此在 4.5 节中，我们采用 0—1 规划方法选取赛 3 场的球队。

4.5 选取赛 3 场的球队的 0—1 规划模型

4.5.1 符号说明

rank 为各队在各自赛区的排名。

$$\text{rank}=(1,2,\cdots,5;1,2,\cdots,5;1,2,\cdots,5)$$

排名依据为上赛季球队在自己分赛区的排名。

4.5.2 设置决策变量

下面以东部联盟为例设置决策变量。西部联盟与东部一致。

x_{ij} 表示第 i 支队和第 j 支队是否比赛，且 $i=1,2,\cdots,15$，分别代表魔术队、奇才队、老鹰队、山猫队、热火队、凯尔特人队、猛龙队、76 人队、篮网队、尼克斯队、活塞队、骑士队、步行者队、公牛队和雄鹿队。$j=1,2,\cdots,15$，分别代表魔术队、奇才队、老鹰队、山猫队、热火队、凯尔特人队、猛龙队、76 人队、篮网队、尼克斯队、活塞队、骑士队、步行者队、公牛队和雄鹿队。

$x_{ij}=1$，表示第 i 支队和第 j 支队打了 3 场比赛。

$x_{ij}=0$，表示第 i 支队和第 j 支队打了 4 场比赛。

4.5.3 目标函数

$$\min=\sum_i\left(\frac{\sum_j x_{ij}\cdot \text{rank}(j)}{4}-3\right)^2$$

$\sum_j x_{ij}\cdot \text{rank}(j)$ 得出各支球队所选打 3 场比赛的球队排名总和；再求平均值，得出所选打 3 场比赛的球队排名的平均排名。由于为赛区排名，只有 1～5 名。因此，平均排名为 3，对每支球队都公平、有利。于是，将求出的平均值和 3 比较。

4.5.4 约束条件

(1) 由于第 i 支队和第 j 支队打比赛，和第 j 支队和第 i 支队打比赛是同一场，所以

$$x_{ij}=x_{ji}\quad(i,j=1,2,\cdots,15)$$

(2) 因为每支球队只能和 4 支球队打 3 场比赛，故

$$\sum_j x_{ij}=4\quad(i=1,2,\cdots,15)$$

$$\sum_i x_{ij}=4\quad(i=1,2,\cdots,15)$$

(3) 由于每支球队与同区的每球队赛 4 场，因此

$$x_{ij}=0\quad(i,j=1,2,\cdots,5)$$

$$x_{ij}=0\quad(i,j=6,7,\cdots,10)$$

$$x_{ij}=0\quad(i,j=11,12,\cdots,15)$$

4.5.5 建立模型

综上所述，建立模型如下：

$$\min = \sum_i \left(\frac{\sum_j x_{ij} \cdot \mathrm{rank}(j)}{4} - 3 \right)^2$$

s. t.

$$x_{ij} = 0,1 \quad (i,j = 1,2,\cdots,15)$$

$$x_{ij} = x_{ij} \quad (i,j = 1,2,\cdots,15)$$

$$\sum_i x_{ij} = 4 \quad (i = 1,2,\cdots,15)$$

$$\sum_j x_{ij} = 4 \quad (i = 1,2,\cdots,15)$$

$$x_{ij} = 0 \quad (i,j = 1,2,\cdots,5)$$

$$x_{ij} = 0 \quad (i,j = 6,7,\cdots,10)$$

$$x_{ij} = 0 \quad (i,j = 11,12,\cdots,15)$$

4.5.6 分配主客场的方法

根据所选 4 队的排名，将最高排名和最低排名列为一组，它们有相同的主客场数（例如，4 级队伍排名分别为 1、2、4、5，则 1 和 5 同时为 2 主 1 客，2 和 4 相反）。如果有相同排名的两队，首先按如上分组，相同排名的球队根据打两次 2 主 1 客和两次 2 客 1 主来权衡分布（注：排名为 1、3、3、5 时，直接依照上述方法，将排名 1、5 的两队列为一组，排名 3、3 的两队列为一组）。

4.5.7 结果表示

将利用 LINGO 软件求解的结果如表 7 所示。

表 7　LINGO 软件求解结果

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1								★			☆		★		☆
2						★		☆	☆					★	
3							☆			★	★			☆	
4								☆			★	☆			★
5						★	☆			★				☆	
6		☆			☆							★		★	
7			★		★							☆	☆		
8	☆	★		★											☆
9		★									☆		☆		★
10			☆		☆							★	★		
11	★		☆	☆					★						
12				★		☆	★			☆					
13	☆						★		★	☆					
14		☆	★		★	☆									
15	★			☆				★	☆						

注：其中，有五角星，代表两队有 3 场比赛；“★”表示 2 主 1 客；“☆”表示 2 客 1 主。

教学方法和教学策略

教学方法	项目教学法		
师生角色	学生为中心;教师是启发者和辅导者		
实施重点	任务驱动,让学生在干中学		
实施步骤	教学过程与内容	教学方法与策略	学时
1. 资讯	(1) 分析并明确工作学习任务 (2) 引导学生搜集相关信息	引导文教学法、讲授法	2
2. 计划	(1) 了解工作计划要素 (2) 制订初步工作计划	小组讨论法 问题引导法 小组总结	1
3. 决策	学生讨论、修改、确定工作计划	师生互动 共同探讨	3
4. 实施	(1) 完成因素指标层次模型 (2) 完成因素层权重计算 (3) 进行各因素指标量化计算 (4) 综合计算总得分,给出评价结果	头脑风暴法 思维导图 小组讨论法	6
5. 检查	对照工作计划及任务要求,检查综合评价方案过程与结果,发现存在的问题与不足,并相应地修改	问题引导法 对比分析法	2
6. 评价	师生共同评价工作任务完成的效果和质量。针对不足,提出改进措施与注意事项	小组讨论 展示	2

成果展示要求

展示成果	1. 工作时间调度模型(学习任务页-zh1)论文(Word 排版) 2. 工作时间调度优化模型建立过程介绍(PPT)
展示方式	1. 每组把促销方案制作成幻灯片进行展示 2. 小组可以选出一个人进行展示,其他同学发现有不完善的,可以补充;也可以分工,每位同学展示方案的一部分
展示顺序	抽签决定
展示时间	每个小组展示时间 7 分钟,回答问题时间 3 分钟

续表

展示要求	工作时间调度模型论文	• 格式正确,排版规范 • 结构完整,有较强的逻辑性和系统性 • 重点突出,包含重要信息和关联点 • 计算准确,论述周详 • 假设合理,推导严密,结论理由充分 • 条理清楚,行文流畅
	PPT 展示	• 格式和视觉效果较好 • 内容完整,条理清楚,逻辑严谨 • 内容、语言和媒体的选择相辅相成 • 体现团队精神 • 语言富有表现力,易于理解 • 有眼神交流,表情生动 • 语速、语调适中

学习评价——个人工作过程报告

学生信息:________________

请在下表中认为描述相符的数字位置画上“√”。

1	小组工作中,个人感觉小组氛围如何?						
	非常好	1	2	3	4	5	非常不好
2	小组成员对本人接受度如何?						
	完全接受	1	2	3	4	5	完全排斥
3	本人的行为是否赢得了小组的理解?						
	非常强	1	2	3	4	5	根本没有
4	在项目中是否学到了东西?						
	非常多	1	2	3	4	5	根本没有
5	对小组工作结果是否满意?						
	非常满意	1	2	3	4	5	完全不满意
6	感觉老师对我们小组工作						
	非常有帮助	1	2	3	4	5	没有帮助和引导

通过下面的调查表,对小组中个人的工作进行评价。

在我的团队中,我能……	评　　价
专业能力 • 按照专业和实际情况展示学习和工作成果 • 独立或在小组中按照时间要求设计工作步骤 • 充分利用专业信息来源和工作辅助工具 • 想出新创意,给出创造性建议 • 区别重要信息和非重要信息	

续表

在我的团队中，我能……	评　　价
方法能力 • 获取、阅读、分析、理解信息资料 • 策划工作步骤，在规定时间内实施 • 根据目标，采用基本的工作技巧 • 组织新观点，对比不同意见和评价 • 采用适当的方法展示自己的观点、想法和结果	
社会能力 • 接受并遵守小组内达成一致的交流和对话准则 • 为自己的观点提出理由和依据 • 能够对不同意见做出适当的反应 • 接受小组分配的任务，并参与任务分配工作 • 正确面对小组内的冲突，并找出解决方法 • 向其他学生提供帮助，向其他学生寻求帮助	
自我认识能力 • 自己检查工作成果是否合理，是否正确 • 自己认识到当前工作的步骤及缺点 • 了解个人在小组中的位置和贡献 • 自己确立合适的工作和行动目标	

学习评价——小组工作评价

每个小组中应该确定一名小组长，作为此次工作的负责人。每名组员应该在各项任务中轮流担任小组长。小组长要注意，是否考虑工作中的每一步，并且实施。当小组工作偏离或背离主题时，应该为小组找到正确的方向。

请小组长在下表右侧各列中，自上而下，将已经完成的工作点用"√"的形式表示出来。

学习任务编号		1	2	3	4	5	6	zh1	zh2
计划阶段	分配任务，弄清任务要求								
	确定工作步骤								
	估计所需时间，制作工作时间表								
实施阶段	顺利完成任务								
	互相帮助，并提供咨询								
	认真、仔细地进行和题目相关的工作								
	定时检查工作进度								
	准时准备好报告								
评价阶段	客观地评价工作结果								
	客观地分析合作情况								
	总结对下次小组工作有用的方法								

对自己和小组其他成员的评价：小组中的每个成员都应在小组工作结束后进行此评价。

是否注意到以下小组规则?	学习任务编号(写出评价分数,5＝极好,0＝极差)							
	1	2	3	4	5	6	zh1	zh2
小组顺利地开展工作								
小组成员对工作任务详细讨论,并确定目标								
小组成员互相帮助并且互相关注								
小组成员听取组员意见并且思考								
小组中的所有成员以同样的工作强度参与工作								
每个组员都记录小组及个人的工作过程								
出现问题时,认真讨论并共同解决								
小组成员专心于主题工作,而没有进行其他活动								
小组讨论中没有出现人身攻击或者谩骂的现象								

参 考 文 献

[1] 姜启源. 数学模型[M]. 4 版. 北京：高等教育出版社，2011.
[2] 边馥萍. 数学模型方法与算法[M]. 北京：高等教育出版社，2005.
[3] 朱道元. 数学建模案例精选[M]. 北京：科学出版社，2003.
[4] 焦云芳. 数学建模入门[M]. 北京：冶金工业出版社，2013.
[5] 杭国英，刘红梅. 计算机应用基础教程主编[M]. 北京：中国铁道出版社，2006.
[6] 延静. 调查技能与分析[M]. 北京：清华大学出版社，2006.
[7] 文杰书院. Excel 2010 公式函数图表与数据分析[M]. 北京：清华大学出版社，2013.
[8] 张志涌. 精通 MATLAB 6.5[M]. 北京：北京航空航天大学出版社，2003.
[9] 袁新生，邵大宏，郝时炼. LINGO 和 Excel 在数学建模中的应用[M]. 北京：科学出版社，2007.
[10] 刘新宪，朱道立. 选择与判断：AHP(层次分析法)[M]. 上海：上海科学普及出版社，1990.
[11] 拉丁. 运筹学——优化模型与算法[M]. 北京：电子工业出版社，2007.
[12] 张立卫. 最优化方法[M]. 北京：科学出版社，2010.
[13] 曾五一. 统计学概论[M]. 北京：首都经济贸易大学出版社，2010.
[14] 蒋绍忠. 数据、模型与决策——基于 Excel 的建模和商务应用[M]. 2 版. 北京：北京大学出版社，2013.
[15] 何晓群，刘文卿. 应用回归分析[M]. 4 版. 北京：中国人民大学出版社，2015.
[16] 李伯德，王国兴. 经济管理数学模型[M]. 北京：科学出版社，2015.
[17] 韩中庚. 数学建模竞赛——获奖论文精选与点评[M]. 北京：科学出版社，2013.
[18] 陈彦光. 地理数学方法[M]. 北京：科学出版社，2011.
[19] 谢金星，薛毅. 优化建模与 LINDO/LINGO 软件[M]. 北京：清华大学出版社，2005.
[20] 刘来福，黄海洋，曾文艺. 数学模型与数学建模[M]. 4 版. 北京：北京师范大学出版社，2014.
[21] 贾俊平，何晓群，金勇进. 统计学[M]. 6 版. 北京：中国人民大学出版社，2015.
[22] 全国大学生数学建模竞赛，http://www.mcm.edu.cn/.